UNIVERSITY OF
CAMBRIDGE

Cambridge Working Papers in Economics

Time Series Models with an EGB2 Conditional Distribution

Michele Caivano and Andrew Harvey

CWPE 1325

Time series models with an EGB2 conditional distribution

Michele Caivano and Andrew Harvey

July 2013

CWPE 1325

Time series models with an EGB2 conditional distribution

Michele Caivano and Andrew Harvey*
Bank of Italy and Faculty of Economics, Cambridge University

July 17, 2013

Abstract

A time series model in which the signal is buried in noise that is non-Gaussian may throw up observations that, when judged by the Gaussian yardstick, are outliers. We describe an observation driven model, based on an exponential generalized beta distribution of the second kind (EGB2), in which the signal is a linear function of past values of the score of the conditional distribution. This specification produces a model that is not only easy to implement, but which also facilitates the development of a comprehensive and relatively straightforward theory for the asymptotic distribution of the maximum likelihood estimator. The model is fitted to US macroeconomic time series and compared with Gaussian and Student-t models. A theory is then developed for an EGARCH model based on the EGB2 distribution and the model is fitted to exchange rate data. Finally dynamic location and scale models are combined and applied to data on the UK rate of inflation.

KEYWORDS: Beta distribution, EGARCH; fat tails; score; robustness; Student's t; Winsorizing.

JEL classification; C22, G17.

*michele.caivano@bancaditalia.it; ach34@cam.ac.uk

1 Introduction

The changing level, or location, of a time series is usually modeled by an ARIMA model or a linear unobserved components model. The statistical treatment of linear Gaussian models is straightforward, with the Kalman filter playing a key role in handling unobserved components. However, time series are often subject to observations that, when judged by the Gaussian yardstick, are outliers. This is a very real issue for financial time series, but it is also relevant for many macroeconomic series as well. Extending time series models to deal with distributions whose tails are heavier than those of the normal is therefore of considerable practical importance.

A new class of models in which the dynamics are driven by the score of the conditional distribution of the observations has recently been developed by Creal et al (2011, 2013) and Harvey (2013). Models of this kind are called dynamic conditional score (DCS) models and they have already proved useful for modeling location and scale. The models are relatively easy to implement and their form facilitates the development of a comprehensive and relatively straightforward theory for the asymptotic distribution of the maximum likelihood estimator.

In an unobserved components signal plus noise model, additive outliers may be captured by letting the noise have a Student t-distribution. The DCS-t model proposed by Harvey and Luati (2012) provides an alternative approach which is observation-driven in that the conditional distribution of the observations is specified. The attractions of using the t-distribution to guard against outliers in static models is well-documented; see, for example, Lange, Little and Taylor (1989). A parametric approach may be compared and contrasted with the methods in the robustness literature; see, for example, Maronna, Martin and Yohai (2006, ch 8). Robust procedures for guarding against additive outliers typically respond to large observations in one of two ways: either the response function converges to a positive (negative) constant for observations tending to plus (or minus) infinity or it goes to zero. These two approaches are usually classified as Winsorizing or as trimming. As is well-known, the score for a t-distribution converges to zero and so can be regarded as a parametric form of trimming. This connection then raises the question as to whether there is a distribution which leads to some form of Winsoring and which is amenable to treatment as a DCS model, as is the Student-t model studied in Harvey and Luati (2012). It turns out that one such distribution is the exponential generalized beta distribution of the second kind (EGB2). The EGB2 distribution was first analyzed in Prentice (1975) and further explored by McDonald and Xu (1995). The aim of this article is to set out the theory for the DCS location model with an EGB2 distribution and to illustrate its practical value. It is worth noting that the need to Winsorize observations is often thought to be necessary in applied work; see Lui, Mitchell and Weale (2011, p 333-4) where the technique is used on UK data.

A DCS-EGB2 model with a dynamic scale may also be developed. This model belongs to the exponential generalized autoregressive heteroskedasticity (EGARCH) class introduced by Nelson (1991) and is complementary to the Student-t model discussed in Creal et al (2011) and Harvey (2013, ch 4). Finally an EGB2 model for time-varying location and scale is developed.

The article is organized as follows. Section 2 reviews a number of issues. The first sub-section sets out a simple unobserved components model and discusses the rationale for letting the dynamics depend on the conditional score. The second sub-section discusses classifications of tail behaviour in distributions and the third shows how these classifications are related to robust estimators for both location and scale. The asymptotic distribution of the maximum likelihood estimator for the first-order DCS model is given in Section 3 and the models for a Student t-distribution and a general error distribution are reviewed. Properties of the EGB2 distribution are reviewed in Section 4 and the DCS location model with an EGB2 distribution is introduced and analysed in Section 5. Section 6 presents some applications to macroeconomic time series and the consequences of a structural shift in the level of a series are briefly discussed in Section 7. The properties of the EGB2-EGARCH model are derived in Section 8 and its use is illustrated with data on exchange rates. Section 9 fits the EGB2 model with time-varying location and scale to UK inflation data and Section 9 concludes.

2 Filters, heavy tails and robust estimation

The first sub-section below sets out a simple unobserved components model and shows how the innovations form of the Kalman filter may be adapted to form a DCS model. The way in which tails of distributions may be classified is reviewed in the second sub-section and finally tail behaviour is related to the considerations of robustness.

2.1 Unobserved components and filters

A simple Gaussian signal plus noise model is

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, \quad \varepsilon_t \sim NID\left(0, \sigma_\varepsilon^2\right), \quad t = 1, \ldots, T, \\ \mu_{t+1} &= \phi \mu_t + \eta_t, \quad \eta_t \sim NID(0, \sigma_\eta^2), \end{aligned} \tag{1}$$

where the irregular and level disturbances, ε_t and η_t respectively, are mutually independent and the notation $NID\left(0, \sigma^2\right)$ denotes normally and independently distributed with mean zero and variance σ^2. The autoregressive parameter is ϕ, while the signal-noise ratio, $q = \sigma_\eta^2/\sigma_\varepsilon^2$, plays the key role in determining how observations should be weighted for prediction and signal extraction. The reduced form (RF) of (1) is an ARMA(1,1) process

$$y_t = \phi y_{t-1} + \xi_t - \theta \xi_{t-1}, \quad \xi_t \sim NID\left(0, \sigma^2\right), \quad t = 1, ..., T \tag{2}$$

but with restrictions on θ. For example, when $\phi = 1$, $0 \leq \theta \leq 1$.

The UC model in (1) is effectively in state space form and, as such, it may be handled by the Kalman filter (KF); see Harvey (1989). The parameters ϕ and q may be estimated by maximum likelihood (ML), with the likelihood function constructed from the one-step ahead prediction errors. The KF can be expressed as a single equation which combines $\mu_{t|t-1}$, the optimal estimator of μ_t based on information at time $t-1$, with y_t in order to produce the best estimator of μ_{t+1}. Writing this equation together with an equation that defines the one-step ahead prediction error, v_t, gives the innovations form of the KF:

$$\begin{aligned} y_t &= \mu_{t|t-1} + v_t, \\ \mu_{t+1|t} &= \phi \mu_{t|t-1} + k_t v_t. \end{aligned} \tag{3}$$

The Kalman gain, k_t, depends on ϕ and q. In the steady-state, k_t is constant. Setting it equal to κ in (3) and re-arranging gives the ARMA model (2) with $\xi_t = v_t$ and $\phi - \kappa = \theta$.

When the noise in (1) comes from a heavy-tailed distribution such as Student's t it can give rise to observations which, when judged against the yardstick of a Gaussian distribution, are additive outliers. As a result fitting a Gaussian model is inefficient and may even yield estimators which are inconsistent. Simulation methods, such as Markov chain Monte Carlo (MCMC) and particle filtering, provide the basis for a direct attack on such non-Gaussian models; see Durbin and Koopman (2012). However, simulation-based estimation can be time-consuming and subject to a degree of uncertainty. In addition the statistical properties of the estimators are not easy to establish.

The DCS approach begins by writing down the distribution of the t-th observation, conditional on past observations. Time-varying parameters are then updated by a suitably defined filter. Such a model is said to be observation driven. In a linear Gaussian UC model, the KF depends on the one step-ahead prediction error. The main ingredient in the DCS filter for non-Gaussian distributions is the replacement of v_t in the KF equation by a variable, u_t, that is proportional to the score of the conditional distribution; compare Maronna, Martin and Yohai (2006, p 272-4) and the references therein. Thus the second equation in (3) becomes

$$\mu_{t+1|t} = \phi\mu_{t|t-1} + \kappa u_t, \tag{4}$$

where κ is treated as an unknown parameter. This filter could be regarded as an approximation to the computer intensive solution for the parameter driven uobserved components model. The attraction of regarding it as a model in its own right is that it becomes possible to derive the asymptotic distribution of the maximum likelihood estimator and generalize in various directions.

2.2 Heavy tails

The Gaussian distribution has kurtosis of three and a distribution is said to exhibit *excess kurtosis,* or to be *leptokurtic,* if its kurtosis is greater than three. Although many researchers take excess kurtosis as defining heavy tails, it is not, in itself, an ideal measure, particularly for asymmetric distributions. Most classifications in the insurance and finance literature begin with the behaviour of the upper tail for a non-negative variable, or one that is only defined above a minimum value; see Asmussen (2003) or Embrechts, Kluppelberg and Mikosch (1997). The two which are relevant here are as follows.

A distribution is said to be *heavy-tailed* if

$$\lim_{y\to\infty} \exp(y/\alpha)\overline{F}(y) = \infty \qquad \text{for all } \alpha > 0, \tag{5}$$

where $\overline{F}(y) = \Pr(Y > y) = 1 - F(y)$ is the survival function. When y has an exponential distribution, $\overline{F}(y) = \exp(-y/\alpha)$, so $\exp(y/\alpha)\overline{F}(y) = 1$ for all y. Thus the exponential distribution is not heavy-tailed.

A distribution is said to be *fat-tailed* if, for a fixed positive value of η,

$$\overline{F}(y) = cL(y)y^{-\eta}, \qquad \eta > 0, \tag{6}$$

where c is a non-negative constant and $L(y)$ is slowly varying[1], that is

$$\lim_{y\to\infty} \frac{L(ky)}{L(y)} = 1.$$

The parameter η is the tail index. The implied PDF is a *power law* PDF

$$f(y) \sim cL(y)\eta y^{-\eta-1}, \qquad y \to \infty, \quad \eta > 0, \tag{7}$$

where $\sim$ is defined such that $a(x) \sim b(x)$ as $x \to x_0$ if $\lim_{x\to x_0}(a/b) \to 1$. The m-th moment exists if $m < \eta$. The Pareto distribution is a simple case in which $\overline{F}(y) = y^{-\eta}$ for $y > 1$. If a distribution is fat-tailed then it must be heavy-tailed, but the converse is not true; see Embrechts, Kluppelberg and Mikosch (1997, p 41-2).

The complement to the Pareto distribution is the power function distribution for which $\overline{F}(y) = y^{\overline{\eta}}$, $0 < y < 1$, $\overline{\eta} > 0$. More generally,

$$\overline{F}(y) = cL(y)y^{\overline{\eta}}, \qquad 0 < y < 1, \quad \overline{\eta} > 0.$$

Hence $f(y) \sim cL(y)\overline{\eta}y^{\overline{\eta}-1}$ as $y \to 0$.

The above criteria are related to the behavior of the conditional score and whether or not it discounts large observations. This, in turn, connects to robustness, as shown in the sub-section following. More specifically, consider a power law PDF, (7), with y divided by a scale parameter, φ, so that $\overline{F}(y/\varphi) = cL(y/\varphi)(y/\varphi)^{-\eta}$ and $f(y) \sim cL(y)\varphi^{-1}\eta(y/\varphi)^{-\eta-1}$. Then

$$\partial \ln f/\partial\varphi \sim \eta/\varphi \quad as \quad y \to \infty \tag{8}$$

and so the score is bounded. With the exponential link function, $\varphi = \exp(\lambda)$,

[1] More generally regularly varying is $\lim_{y\to\infty}(L(ky)/L(y)) = k^{\beta}$; see Embrechts, Kluppelberg and Mikosch (1997, p. 37, 564). Fat-tailed distributions are regularly varying with $\eta = -\beta > 0$.

$\partial \ln f/\partial\lambda \sim \eta$ as $y \to \infty$. Similarly as $y \to 0$, $\partial \ln f/\partial\lambda \sim \overline{\eta}$.

The logarithm of a variable with a fat-tailed distribution has exponential tails. Let x denote a variable with a fat-tailed distribution in which the scale is written as $\varphi = \exp(\mu)$ and let $y = \ln x$. Then for large y

$$f(y) \sim cL(e^y)\eta e^{-\eta(y-\mu)}, \qquad \eta > 0, \quad as \quad y \to \infty,$$

whereas as $y \to -\infty$, $f(y) \sim cL(e^y)\overline{\eta}e^{\overline{\eta}(y-\mu)}$, $\overline{\eta} > 0$. Thus y is not heavy-tailed, but it may exhibit excess kurtosis. The score with respect to location, μ, is the same as the original score with respect to the logarithm of scale and so tends to η as $y \to \infty$.

2.3 Robust estimation

The location-dispersion model is

$$y_t = \mu + \varphi\varepsilon_t, \qquad t = 1, ..., T, \tag{9}$$

where ε_t is a standardized variable with PDF $\rho(\varepsilon_t)$ and the scale, φ, for the variable $y_t - \mu$ is called the dispersion for y_t; see Maronna, Martin and Yohai (2006, p37-8). The density for y_t is

$$f(y_t; \mu, \varphi, \xi) = \varphi^{-1} f((y_t - \mu)/\varphi),$$

where ξ denotes one or more shape parameters, and the scores for μ and φ are given by differentiating $\ln f(y_t) = \rho((y_t - \mu)/\varphi) - \ln\varphi$. The score for location is

$$\frac{\partial \ln f_t}{\partial\mu} = \frac{\partial\rho(z_t)}{\partial\mu} = \psi_L(z_t),$$

where $z_t = (y_t - \mu)/\varphi$, whereas the score for scale is

$$\frac{\partial \ln f_t}{\partial\varphi} = \frac{\partial\rho(z_t)}{\partial\varphi} - \frac{1}{\varphi} = \psi_S(z_t) - \frac{1}{\varphi}.$$

Note that

$$\psi_S(z_t) = \varphi^{-1} z_t \psi_L(z_t). \tag{10}$$

If the scale is parameterized with an exponential link function, $\varphi = \exp \lambda$, the score is

$$\frac{\partial \ln f}{\partial \lambda} = \frac{\partial \rho(z_t)}{\partial \lambda} - 1 = \psi_S(z_t) - 1 = z_t \psi_L(z_t) - 1. \tag{11}$$

Remark 1 *If $x_t = \varphi \varepsilon_t$ in (9) and $y_t = \ln x_t$, then the logarithm of the scale parameter for x_t, that is $\ln \varphi$, becomes the location for y_t. Hence $\psi_L(y_t) = \psi_S(x_t)$.*

The ML estimators are asymptotically efficient, assuming certain regularity conditions hold. More generally $\rho(.)$ may be any function deemed to yield estimators with good statistical properties. In particular, the estimators should be robust to observations which would be considered to be outliers for a normal distribution. When normality is assumed, the ML estimators of the mean and variance are just the corresponding sample moments, but these can be subject to considerable distortion when outliers are present. Robust estimators, on the other hand, are resistant to outliers while retaining relatively high efficiency when the data are from a normal distribution.

The M-estimator, which features prominently in the robustness literature, has a Gaussian response until a certain threshold, K, whereupon it is constant; see Maronna, Martin and Yohai (2006, p 25-31). This is known as Winsorizing as opposed to trimming, where observations greater than K in absolute value are given a weight of zero[2].

3 DCS location models

The stationary first-order DCS model corresponds to the Gaussian innovations form, (3), and is

$$\begin{aligned} y_t &= \mu_{t|t-1} + v_t = \mu_{t|t-1} + \exp(\lambda)\varepsilon_t, \qquad t = 1, ..., T, \\ \mu_{t+1|t} &= \delta + \phi \mu_{t|t-1} + \kappa u_t, \qquad |\phi| < 1, \end{aligned} \tag{12}$$

where $\omega = \delta/(1-\phi)$ is the unconditional mean of $\mu_{t|t-1}$, ε_t is a serially independent, standardized variate and u_t is proportional to the conditional score, that is $u_t = k.\partial \ln f(y_t \mid y_{t-1}, y_{t-2}, ...)/\partial \mu_{t|t-1}$, where k is a constant.

[2] In both cases a (robust) estimate of scale needs to be pre-computed and the process of computing M-estimates is then often iterated to convergence. The same is true for other $\psi-$functions in the literature.

More generally, an ARMA-type model of order (p, r) is

$$\mu_{t+1|t} = \delta + \phi_1 \mu_{t|t-1} + ... + \phi_p \mu_{t-p+1|t-p} + \kappa_0 u_t + \kappa_1 u_{t-1} + ... + \kappa_r u_{t-r}. \quad (13)$$

In the Gaussian case $u_t = y_t - \mu_{t|t-1}$ and if q is defined as $max(p, r+1)$, then y_t is an $ARMA(p, q)$ process with MA coefficients $\theta_i = \phi_i - \kappa_{i-1}$, $i = 1, .., q$. Nonstationary ARIMA-type models may also be constructed as may structural times series models with trend and seasonal components. Explanatory variables can be introduced into DCS models, as described in Harvey and Luati (2012).

Maronna, Martin and Yohai (2006, sect 8.6 and 8.8) give a robust algorithm for AR and ARMA models with additive outliers. For a first-order model their filter is essentially the same as (12) except that their dynamic equation is driven by a robust $\psi - function$ and they regard the model as an approximation to a UC model[3].

3.1 Estimation

The ML estimates of the parameters, $\boldsymbol{\psi} = (\kappa, \phi, \omega)'$, in a DCS location model can be obtained by maximizing the log-likelihood function with respect to the unknown parameters. The asymptotic distribution of the ML estimator in a first-order DCS model is derived in Harvey (2013). Define

$$\begin{aligned} a &= \phi + \kappa E\left(\frac{\partial u_t}{\partial \mu}\right) \quad (14) \\ b &= \phi^2 + 2\phi\kappa E\left(\frac{\partial u_t}{\partial \mu}\right) + \kappa^2 E\left(\frac{\partial u_t}{\partial \mu}\right)^2 \geq 0 \\ c &= \kappa E\left(u_t \frac{\partial u_t}{\partial \mu}\right), \end{aligned}$$

[3] Muler, Peña and Yohai (2009, p817) note two shortcomings of the estimates obtained in this way. They write: 'First, these estimates are asymptotically biased. Second, there is not an asymptotic theory for these estimators, and therefore inference procedures like tests or confidence regions are not available.' They then suggest a different approach and show that it allows an asymptotic theory to be developed. As with robust estimation generally, a preliminary robust estimate of scale is needed.

where unconditional and conditional expectations are the same. When scale and shape parameters are known and $b<1$, the information matrix for a single observation is time-invariant and given by

$$\mathbf{I}(\boldsymbol{\psi})=(\sigma_u^2/k^2)\mathbf{D}(\boldsymbol{\psi}), \tag{15}$$

where σ_u^2/k^2 is the information quantity for a single observation and

$$\mathbf{D}(\boldsymbol{\psi})=\mathbf{D}\left(\begin{array}{c}\widetilde{\kappa}\\ \widetilde{\phi}\\ \widetilde{\omega}\end{array}\right)=\frac{1}{1-b}\left[\begin{array}{ccc}A & D & E\\ D & B & F\\ E & F & C\end{array}\right], \tag{16}$$

with

$$\begin{aligned} A &= \sigma_u^2, \qquad B=\frac{\kappa^2\sigma_u^2(1+a\phi)}{(1-\phi^2)(1-a\phi)}, \qquad C=\frac{(1-\phi)^2(1+a)}{1-a},\\ D &= \frac{a\kappa\sigma_u^2}{1-a\phi}, \qquad E=c(1-\phi)/(1-a) \qquad \textit{and} \qquad F=\frac{ac\kappa(1-\phi)}{(1-a)(1-a\phi)}. \end{aligned}$$

The ML estimator is asymptotically normal with covariance matrix given by the inverse of (15). In a location model, $c=0$ when the distribution is symmetric.

More generally consider a vector of parameters, $\boldsymbol{\theta}$, such that $\boldsymbol{\theta}=(\theta_1,\boldsymbol{\theta}_2')'$. Suppose that $\boldsymbol{\theta}_2$ consists of $n-1\geq 1$ fixed parameters, whereas θ_1 is time-varying and depends on a set of parameters, $\boldsymbol{\psi}$. When the terms in the information matrix of the static model that involve θ_1, including cross-products, do not depend on θ_1,

$$\mathbf{I}\left(\begin{array}{c}\boldsymbol{\psi}\\ \boldsymbol{\theta}_2\end{array}\right)=\left[\begin{array}{cc}E\left(\frac{\partial\ln f_t}{\partial\theta_1}\right)^2\mathbf{D}(\boldsymbol{\psi}) & \mathbf{d}E\left(\frac{\partial\ln f_t}{\partial\theta_1}\frac{\partial\ln f_t}{\partial\boldsymbol{\theta}_2'}\right)\\ E\left(\frac{\partial\ln f_t}{\partial\theta_1}\frac{\partial\ln f_t}{\partial\boldsymbol{\theta}_2}\right)\mathbf{d}' & E\left(\frac{\partial\ln f_t}{\partial\boldsymbol{\theta}_2}\frac{\partial\ln f_t}{\partial\boldsymbol{\theta}_2'}\right)\end{array}\right], \tag{17}$$

where $\mathbf{D}(\boldsymbol{\psi})$ is the matrix in (16) and $\mathbf{d}=(0,0,(1-\phi)/(1-a))'$.

3.2 Dynamic Student-t location model

The t_ν-distribution with a location of μ and scale of φ has probability density function (PDF)

$$f\left(y;\mu,\varphi,\nu\right)=\frac{\Gamma\left(\left(\nu+1\right)/2\right)}{\Gamma\left(\nu/2\right)\varphi\sqrt{\pi\nu}}\left(1+\frac{\left(y-\mu\right)^2}{\nu\varphi^2}\right)^{-(\nu+1)/2},\qquad \varphi,\nu>0, \tag{18}$$

where ν is the degrees of freedom and $\Gamma\left(.\right)$ is the gamma function. Moments exist only up to and including $\nu-1$. For $\nu>2$, the variance is $\sigma^2=\left\{\nu/\left(\nu-2\right)\right\}\varphi^2$. The excess kurtosis, that is the amount by which the normal distribution's kurtosis of three is exceeded, is $6/(\nu-4)$, provided that $\nu>4$. The t_ν distribution has fat tails for finite ν with the tail index given by ν; see McNeil et al (2005, p293).

When the location changes over time, it may be captured by a model in which, conditional on past observations, y_t has a t_ν-distribution with $\mu_{t|t-1}$ generated by a linear function of

$$u_t=\left(1+\nu^{-1}e^{-2\lambda}(y_t-\mu_{t|t-1})^2\right)^{-1}v_t,\quad t=1,...,T, \tag{19}$$

where $v_t=y_t-\mu_{t|t-1}$ is the prediction error. Differentiating the log-density shows that u_t is proportional to the conditional score, $\partial\ln f_t/\partial\mu_{t|t-1}=(\nu+1)\nu^{-1}\exp(-2\lambda)u_t$. For low degrees of freedom, the score function is such that observations that would be seen as outliers for a Gaussian distribution are far less influential. As $|y|\to\infty$, the response tends to zero. Redescending M-estimators, which feature in the robustness literature, have the same property. For example, Tukey's biweight function is $\psi(z)=[1-(x/4.685)^2]_+^2z$, where $[\ ]_+$ denotes the positive part of the term in $[\ ]$.This function implements 'soft trimming', as opposed to metric trimming, where $\psi(z)=z$ for $|z|\leq K$ and is zero thereafter. The t score function is even softer.

The variable u_t can be written $u_t=(1-b_t)(y_t-\mu_{t|t-1})$, where

$$b_t=\frac{(y_t-\mu_{t|t-1})^2/\nu\exp(2\lambda)}{1+(y_t-\mu_{t|t-1})^2/\nu\exp(2\lambda)},\qquad 0\leq b_t\leq 1,\quad 0<\nu<\infty, \tag{20}$$

is distributed as $beta(1/2,\nu/2)$; see Harvey (2013, Chapter 3). The $u_t's$ are $IID(0,\sigma_u^2)$ and symmetrically distributed with $\sigma_u^2=\nu^2(\nu+3)^{-1}(\nu+1)^{-1}\exp(2\lambda)$. Since the $u_t's$ are $IID(0,\sigma_u^2)$, $\mu_{t|t-1}$ is weakly and strictly stationary so long as $|\phi| < 1$. Although determining the statistical properties of $\mu_{t|t-1}$ requires assuming that it started in the infinite past, the filter needs to be initialized in practice and this may be done by setting $\mu_{1|0} = \omega$.

All moments of u_t exist and the existence of moments of y_t is not affected by the dynamics. The autocorrelations can be found from the infinite MA representation; the patterns are as they would be for a Gaussian model. For a stationary first-order model, as in (12), an analytic expression for the information is given in Harvey (2013).

There are a number of ways in which skewness may be introduced into a t-distribution. One possibility is the method proposed by Fernandez and Steel (1998). There is a minor technical issue in that the score is not differentiable at the mode but as, Zhu and Galbraith (2011) show, the asymptotic theory for the ML estimator still goes through in the usual way. The asymptotic theory for the DCS skew-t location model also goes through; see Harvey (2013, Section 3.11).

3.3 General error distribution

The PDF of the general error distribution, denoted $GED(\upsilon)$, is

$$f\left(y;\mu,\varphi,\upsilon\right) = \left[2^{1+1/\upsilon}\varphi\Gamma(1+1/\upsilon)\right]^{-1}\exp(-\left|(y-\mu)/\varphi\right|^{\upsilon}/2), \qquad \varphi,\upsilon > 0,$$

where φ is a scale parameter, related to the standard deviation by the formula

$$\sigma = 2^{1/\upsilon}(\Gamma\left(3/\upsilon\right)/\Gamma\left(1/\upsilon\right))^{1/2}\varphi.$$

The normal distribution is obtained when $\upsilon = 2$, in which case $\sigma = \varphi$. Setting $\upsilon = 1$ gives the Laplace, or double exponential, distribution, for which in which case $\sigma = 2\sqrt{2}\varphi$. Therefore when $1 \leq \upsilon \leq 2$, which is usually the case, the GED distribution provides a continuum between the normal and Laplace. The kurtosis is $\Gamma\left(5/\upsilon\right)\Gamma\left(1/\upsilon\right)/\Gamma\left(3/\upsilon\right)$, so for $\upsilon = 1$ the excess kurtosis is three.

The score for μ is

$$\frac{\partial \ln f_t(\mu,\lambda,\upsilon)}{\partial\mu} = \frac{\upsilon}{2\varphi^{\upsilon}} sgn(y-\mu)\left|y_t-\mu\right|^{\upsilon-1}, \quad t = 1,...,T,$$

but it is convenient to define u_t in the DCS model as $(2\varphi/\upsilon)$ times the score. For $\upsilon = 1$, the score is $sgn(y_t - \mu_{t|t-1})$, except at $y_t = \mu_{t|t-1}$, where it is not defined. The value of σ_u^2 is unity for both $\upsilon = 2$ and $\upsilon = 1$ because $\sigma_u^2 = 2^{2(\upsilon-1)/\upsilon}\Gamma(2-\upsilon^{-1})/\Gamma(1/\upsilon)$.

The ML estimator of μ when $\upsilon = 1$ is the median. The usual asymptotic properties of the ML estimator of μ can be shown to hold, though the proof is non-standard for $\upsilon < 2$ because the score function is not continuous at $y = \mu$; see Zhu and Zinde-Walsh (2009). However, for the DCS model, the asymptotic theory in sub-section 3.1 runs into difficulties when $\upsilon \leq 1.5$ because the higher order moments upon which b depends do not exist. Specifically

$$\frac{\partial u_t}{\partial \mu_{t|t-1}} = (\upsilon - 1)\left|\frac{y_t - \mu_{t|t-1}}{\exp(\lambda)}\right|^{\upsilon-2},$$

and the expectation of its square only exists fo $\upsilon > 1.5$. The ability of the model to capture leptokurtic behaviour is therefore limited because the excess kurtosis for $\upsilon = 1.5$ is only 0.762.

4 Exponential generalized beta distribution

The exponential generalized beta distribution of the second kind (EGB2) is obtained by taking the logarithm of a variable with a GB2 distribution. The PDF of a GB2$(\alpha, \nu, \xi, \varsigma)$ variate is

$$f(x) = \frac{\nu(x/\alpha)^{\nu\xi-1}}{\alpha B(\xi,\varsigma)\left[(x/\alpha)^{\nu}+1\right]^{\xi+\varsigma}}, \qquad \alpha, \nu, \xi, \varsigma > 0, \tag{21}$$

where α is the scale parameter, ν, ξ and ς are shape parameters and $B(\xi, \varsigma)$ is the beta function; see Kleiber[4] and Kotz (2003, ch6). The GB2 distribution contains many important distributions as special cases, including the Burr ($\xi = 1$) and log-logistic ($\xi = 1, \varsigma = 1$). GB2 distributions are fat tailed for finite ξ and ς with upper and lower tail indices of $\eta = \varsigma\nu$ and $\overline{\eta} = \xi\nu$ respectively. The absolute value[5] of a t_f variate is GB2$(f^{1/2}\varphi, 2, 1/2, f/2)$ with tail index is $\eta = \overline{\eta} = f$.

[4] Note that Kleiber and Kotz (2003) have α and ν in reverse order, ie they write $GB2(\nu, \alpha, \xi, \varsigma)$. Our preference is to put α first because it becomes a location parameter in EGB2.

[5] On the other hand the absolute value of GED variate is generalized gamma with the product of the scale parameters equal to one; in the notation of Harvey (2013, ch 5), $\gamma\upsilon = 1$.

The properties which EGB2 inherits from GB2 have important implications for the score function and hence for robustness to additive outliers. The connection between the score for a t-distribution and redescending M-estimators is well-known; see sub-section 3.3 and Maronna, Martin and Yohai (2006, pp 29). The fact that the EGB2 distribution gives a gentle form of Winsorizing is less well-known, although its robustness properties have been studied by McDonald and White (1993).

If x is distributed as GB2$(\alpha, \nu, \xi, \varsigma)$ and $y = \ln x$, the PDF of the EGB2 variate y is

$$f(y; \mu, \nu, \xi, \varsigma) = \frac{\nu \exp\{\xi(y-\mu)\nu\}}{B(\xi, \varsigma)(1 + \exp\{(y-\mu)\nu\})^{\xi+\varsigma}}. \tag{22}$$

What was the logarithm of scale in GB2 now becomes location in EGB, that is $\ln \alpha$ becomes μ. Furthermore ν is now a scale parameter, but ξ and ς are still shape parameters and they determine skewness and kurtosis.

The EGB2 distribution has exponential tails; see the end of sub-section 2.2. All moments exist, the first four being as follows:

$$\text{Mean: } E(y) = \mu + \nu^{-1}[\psi(\xi) - \psi(\varsigma)] \tag{23}$$

$$\text{Variance: } \sigma^2 = E(y - E(y))^2 = \nu^{-2}[\psi'(\xi) + \psi'(\varsigma)] \tag{24}$$

$$\text{Skewness: } \frac{E(y - E(y))^3}{\sigma^3} = \frac{\psi''(\xi) - \psi''(\varsigma)}{[\psi'(\xi) + \psi'(\varsigma)]^{3/2}} \tag{25}$$

$$\text{Kurtosis: } \frac{E(y - E(y))^4}{\sigma^4} = \frac{\psi'''(\xi) + \psi'''(\varsigma)}{[\psi'(\xi) + \psi'(\varsigma)]^2} + 3, \tag{26}$$

where ψ, ψ', ψ'' and ψ''' are polygamma functions of order 0, 1, 2 and 3 respectively. The EGB2 distribution is positively (negatively) skewed when $\xi > \varsigma$ ($\xi < \varsigma$) and its kurtosis decreases as ξ and ς increase. Skewness ranges between -2 and 2 and kurtosis[6] lies between 3 and 9. There is excess kurtosis for finite ξ and/or ς.

Although ν is a scale parameter, it is the inverse of what would be considered a more conventional measure of scale. Thus scale is better defined as

[6] The maximum kurtosis in the symmetric case is 6 and is for $\xi = \varsigma = 0$. The kurtosis of 9 is achieved when ξ(or ς) $= 0$ and ς(or ξ)$=\infty$.

$1/\nu$ or as the standard deviation[7]

$$\sigma = \sqrt{\psi'(\xi) + \psi'(\varsigma)}/\nu = h(\xi, \varsigma)/\nu = h/\nu. \tag{27}$$

The following results will be used in a number of places when the σ parameterization is adopted.

Lemma 1 *Let $\xi = \varsigma$. Then (i) $\xi h^2 = 2$ as $\xi \to \infty$, and $\xi h \to 2/h \to \infty$.(ii) $\xi h = \sqrt{2}$ for $\xi = 0$. Or equivalently, (i) $\xi\psi'(\xi) = 1$ and $\xi\sqrt{\psi'(\xi)} \to 1/\sqrt{\psi'(\xi)}$ as $\xi \to \infty$ (ii) $\xi\sqrt{\psi'(\xi)} = 1$ for $\xi = 0$.*

The PDF of the EGB2 with the standard deviation parameterization is

$$f(y; \mu, \sigma, \xi, \varsigma) = \frac{h \exp\{\xi h(y-\mu)/\sigma\}}{\sigma B(\xi, \varsigma)(1 + \exp\{h(y-\mu)/\sigma\})^{\xi+\varsigma}}.$$

When $\xi = \varsigma$, the EGB2 distribution is symmetric; for $\xi = \varsigma = 1$ it is a logistic distribution and when $\xi = \varsigma \to \infty$ it tends to a normal distribution. The case of $\xi = \varsigma = 0$ is important, but rarely mentioned in the literature.

Proposition 1 *When $\xi = \varsigma = 0$ in the EGB2 of (27), the distribution is double exponential or Laplace.*

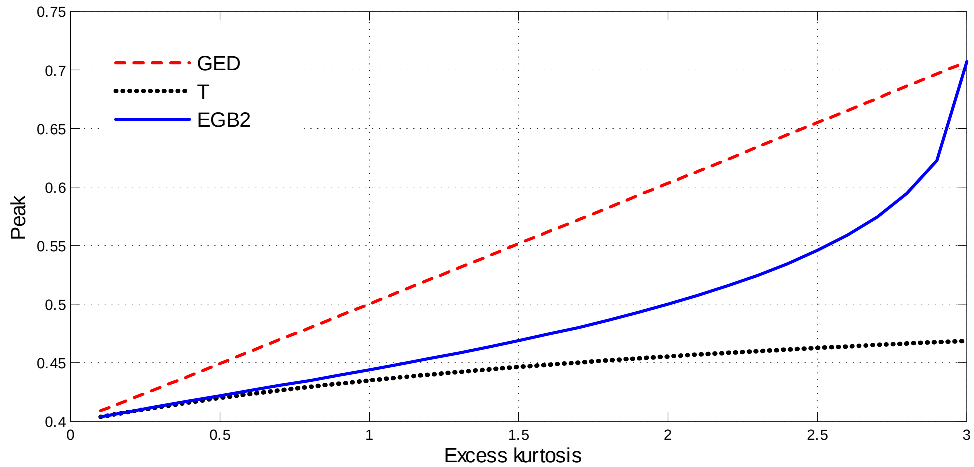
Proof. For simplicity of notation let $\mu = 0$. Suppose $y \leq 0$. Then, noting that $\Gamma(kz)/\Gamma(z) = 1/k$ and writing $\Gamma(\xi) = \xi^{-1}\Gamma(\xi+1)$,

$$\begin{aligned} f(y; 0, h/\sigma, \xi, \xi) &= \frac{h\Gamma(2\xi)\xi \exp\{-\xi h\,|y/\sigma|\}}{\sigma\Gamma(\xi)\Gamma(\xi)(1 + \exp\{-\xi h\,|y/\sigma|\})^{2\xi}} \\ &= \frac{\xi h\Gamma(2\xi+1)\xi \exp\{-\xi h\,|y/\sigma|\}}{\sigma\Gamma(\xi+1)\Gamma(\xi+1)2\xi(1 + \exp\{-\xi h\,|y/\sigma|\})^{2\xi}} \end{aligned}$$

Cancelling the $\xi's$, setting $\xi = 0$ and noting that $\xi h = \sqrt{2}$ when $\xi = 0$ gives the result because $\sigma = 2\sqrt{2}\varphi$. When $y > 0$ we first need to multiply numerator and denominator by $\exp\{-2\xi hy/\sigma\}$ before invoking the same argument. ■

Plots of the (symmetric) EGB2, GED and Student's t with the same excess kurtosis shows them to be very similar. It is difficult to see the heavier

[7] With the standard deviation parameterization, $E(y) = \mu + \sigma[\psi(\xi) - \psi(\varsigma)]/h$.



tails of the t distribution, and the only discernible difference among the three distributions is in the peak, which is higher and more pointed for the GED. The EGB2 in turn is more peaked than the t. As the excess kurtosis increases, the differences between the peaks become more marked; see Figure **??**.

5 Dynamic EGB2 location model

The score function for the GB2 distribution with respect to $\ln \alpha$ is bounded, reflecting the fact that the distribution has a fat tail and α is a scale parameter; see sub-section 2.3. The score function with respect to location in the EGB2 distribution is of exactly the same form and so inherits the boundedness property; see Remark 1. Specifically,

$$\ln f_t(y_t;\nu,\xi,\varsigma) = \ln \nu - \nu\xi\mu_{t|t-1} + \nu\xi y_t - (\xi+\varsigma)\ln((e^{y_t-\mu_{t|t-1}})^{\nu}+1) - \ln B(\xi,\varsigma),$$

with $\mu_{t|t-1}$ as in (12), and so

$$\frac{\partial \ln f_t}{\partial \mu_{t|t-1}} = \nu(\xi+\varsigma)b_t(\xi,\varsigma) - \nu\xi, \qquad t = 1, ..., T,$$

where

$$b_t(\xi,\varsigma) = \frac{e^{(y_t-\mu_{t|t-1})\nu}}{e^{(y_t-\mu_{t|t-1})\nu}+1}.$$

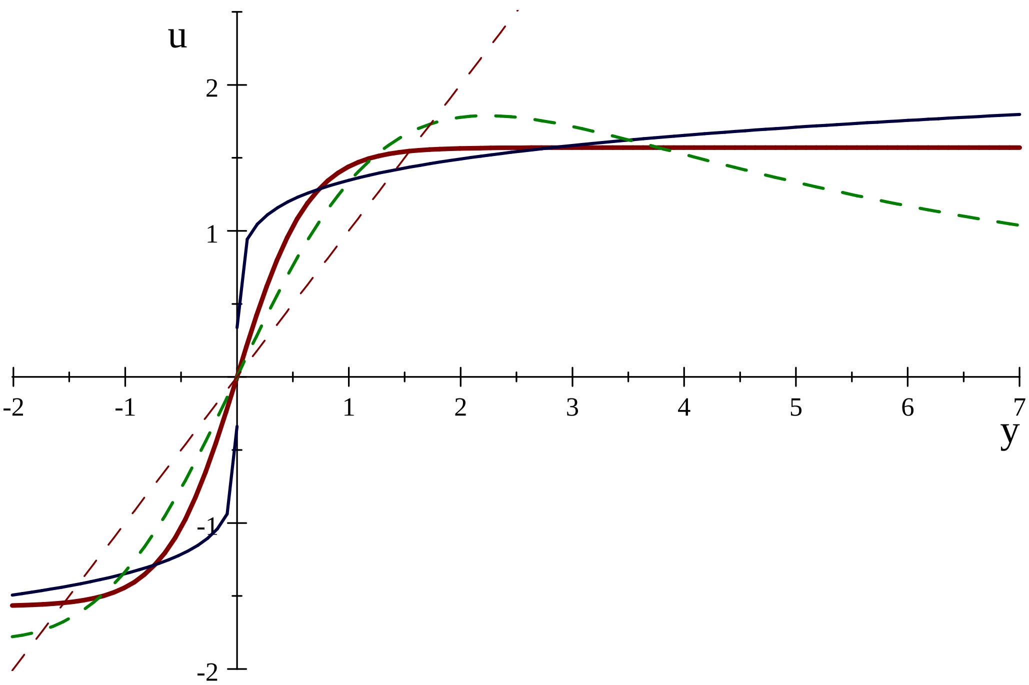


Figure 1: Score functions for EGB2 (thick line), GED (medium line) and t (thick dash), all with excess kurtosis of 2. Thin line shows normal score. (Note that $\sigma = 1$ and $u(-y) = -u(y)$ for $y > 0$).

Because $0 \leq b_t(\xi, \varsigma) \leq 1$, it follows that as $y \rightarrow \infty$, the score approaches an upper bound of $\nu\varsigma$, whereas $y \rightarrow -\infty$ gives a lower bound of $\nu\xi$.

It will prove more convenient to replace ν by h/σ and to define u_t as

$$u_t = \sigma^2 \frac{\partial \ln f_t}{\partial \mu_{t|t-1}} = \sigma h[(\xi + \varsigma)b_t(\xi, \varsigma) - \xi]. \tag{28}$$

We note that the upper and lower bounds are $\sigma\sqrt{2}$ and $-\sigma\sqrt{2}$ respectively when $\varsigma = \xi = 0$. On the other hand, there is no upper (lower) bound for $\varsigma (or$ $\xi) \rightarrow \infty$ because $h\varsigma \rightarrow \infty$ (as does $h\xi$). As $\varsigma = \xi \rightarrow \infty$, the distribution becomes normal and so for large ς and ξ, $u_t \simeq y_t - \mu_{t|t-1}$.

Figure 1 shows the score functions for standardized ($\sigma = 1$) EGB2, GED and t distributions, all with excess kurtosis of two. The shape parameters for the three distributions are $\xi = 0.5$, $\upsilon = 1.148$ and $\nu = 7$. Given the apparent similarity of the PDFs, the difference in the behaviour of the score functions is striking. The score for the t distribution is redescending, reflecting the fact that it has fat tails, as defined in sub-section 2.2. There is no upper bound with GED, except when it becomes a Laplace distribution and the score is $\pm\sqrt{2}$ for $y \neq 0$. Neither the EGB2 nor the GED distribution has heavy tails. However, the EGB2 distribution has exponential tails, whereas the GED distribution is super-exponential for $\upsilon > 1$. Hence the EGB2 score is bounded and what we get is a gentle form of Winsorizing.

Andres and Harvey (2012) study time-varying scale in a GB2 DCS model parameterized with an exponential link function. Many of the results given there and in Harvey (2013) therefore apply to the EGB2 model with time-varying location. In particular, the variable $b_t(\xi, \varsigma)$ is IID with a $beta(\xi, \varsigma)$ distribution at the true parameter values. Thus, it is easy to confirm that $E(u_t) = 0$ and to obtain the variance of u_t as

$$\sigma_u^2 = \sigma^2 h^2 \xi\varsigma/(\xi + \varsigma + 1).$$

As $\xi, \varsigma \to \infty$, $\sigma_u^2 \to \sigma^2$.

5.1 Maximum likelihood estimation

The asymptotic distribution of the ML estimators of the parameters in a dynamic location model with an EGB2 distribution is the same as for a dynamic model for the logarithm of scale with a GB2 distribution. Hence the information matrix is as in Harvey (2013, ch 5). However, an exponential link function for scale ensures that it remains positive and it is also crucial for the development of an asymptotic theory when the scale is allowed to be time-varying later in the paper. For many puposes, it is more convenient to parameterize the scale in terms of the standard deviation in (27) and so ν is replaced by $h \exp(-\lambda_\sigma)$. Unfortunately, the presence of $h = h(\xi, \varsigma)$ complicates the information matrix, as shown in the appendix. Thus it is simpler to just replace ν by $\exp(-\lambda)$, where $\lambda = \lambda_\sigma - \ln h$, if asymptotic standard errors are to be computed. The likelihood function can still be maximized with respect to λ_σ and, in fact, this turns out to be much better for stability and convergence of the numerical optimization. Standard errors are of little practical importance for scale parameters and the standard errors of the other parameters do not depend on the parameterization of the scale.

Proposition 2 *Define*

$$a = \phi - \kappa \frac{\xi\varsigma}{\xi+\varsigma+1}, \qquad c = \kappa \frac{e^{-\lambda}\varsigma\xi(\xi-\varsigma)}{(\varsigma+\xi+2)(\varsigma+\xi+1)},$$

$$b = \phi^2 - 2\phi\kappa \frac{\xi\varsigma}{\xi+\varsigma+1} + \kappa^2 \frac{(\xi+\varsigma)\xi\varsigma\,(\varsigma+1)\,(\xi+1)}{(\varsigma+\xi+3)(\varsigma+\xi+2)(\varsigma+\xi+1)} \tag{29}$$

and note that $k = \sigma^2$. *For a conditional EGB2 distribution, with* ν *replaced by* $\exp(-\lambda)$, *and a first-order stationary dynamic model for the location with* $b < 1$, *the limiting distribution of* $\sqrt{T}(\widetilde{\boldsymbol{\psi}}' - \boldsymbol{\psi}', \widetilde{\lambda} - \lambda, \widetilde{\xi} - \xi, \widetilde{\varsigma} - \varsigma)'$ *is multivariate normal with covariance matrix given by the inverse of*

$$\mathbf{I}\begin{pmatrix} \boldsymbol{\psi} \\ \lambda \\ \xi \\ \varsigma \end{pmatrix} = \begin{bmatrix} \frac{e^{-2\lambda}\xi\varsigma}{1+\xi+\varsigma}\mathbf{D}(\boldsymbol{\psi}) & I_{12}\mathbf{d} & \frac{e^{-\lambda}\varsigma}{\xi+\varsigma}\mathbf{d} & \frac{-e^{-\lambda}\xi}{\xi+\varsigma}\mathbf{d} \\ I_{21}\mathbf{d}' & I_{22} & I_{23} & I_{24} \\ \frac{e^{-\lambda}\varsigma}{\xi+\varsigma}\mathbf{d}' & I_{23} & \psi'(\xi) - \psi'(\xi+\varsigma) & -\psi'(\xi+\varsigma) \\ \frac{-e^{-\lambda}\xi}{\xi+\varsigma}\mathbf{d}' & I_{24} & -\psi'(\xi+\varsigma) & \psi'(\varsigma) - \psi'(\xi+\varsigma) \end{bmatrix}$$

with

$$I_{21} = I_{12} = \frac{-e^{-\lambda}(\xi - \varsigma - \xi\varsigma\,(\psi(\xi) - \psi(\varsigma)))}{1+\xi+\varsigma}, \qquad I_{23} = I_{32} = \frac{\varsigma\,(\psi(\xi) - \psi(\varsigma)) - 1}{\xi+\varsigma},$$

$$I_{24} = I_{42} = \frac{\xi\,(\psi(\varsigma) - \psi(\xi)) - 1}{\xi+\varsigma}$$

and

$$I_{22} = \frac{\xi\varsigma}{1+\xi+\varsigma}\left[\left(\psi'(\xi) + \psi'(\varsigma) + \left(\psi(\varsigma) - \psi(\xi) + \frac{\xi-\varsigma}{\xi\varsigma}\right)^2\right) - \left(\frac{\xi^2+\varsigma^2}{\xi^2\varsigma^2}\right)\right] + 1.$$

Proof. Using (14) to evaluate a, b and c gives

$$E\left[\frac{\partial u_t}{\partial \mu}\right] = -(\xi+\varsigma)E(b_t(1-b_t)) = \frac{-\xi\varsigma}{\xi+\varsigma+1},$$

$$E\left[\frac{\partial u_t}{\partial \mu}\right]^2 = (\xi+\varsigma)^2 E[b_t^2(1-b_t)^2] = \frac{(\xi+\varsigma)\xi\varsigma\,(\varsigma+1)\,(\xi+1)}{(\varsigma+\xi+3)(\varsigma+\xi+2)(\varsigma+\xi+1)}$$

and

$$E\left[u_t \frac{\partial u_t}{\partial \mu}\right] = -(\xi+\varsigma)^2 E[b_t^2(1-b_t)] + (\xi+\varsigma)E[b_t(1-b_t)]$$
$$= \frac{\xi^2\varsigma}{(\xi+\varsigma+1)} - \frac{\xi\varsigma(\xi+\varsigma)(\xi+1)}{(\varsigma+\xi+2)(\varsigma+\xi+1)}.$$

■

Because u_t is the score times the variance, the expressions for a, b and c in (29) are independent of scale.

Note that if the estimator of λ_σ has been computed, $\lambda = \lambda_\sigma - \ln h$.

5.2 Estimation for a symmetric distribution

When ξ and ς are constrained to be equal, the information matrix is

$$\mathbf{I}\begin{pmatrix} \boldsymbol{\psi} \\ \lambda \\ \xi \end{pmatrix} = \begin{bmatrix} \frac{e^{-2\lambda}\xi^2}{1+2\xi}\mathbf{D}(\boldsymbol{\psi}) & 0 & 0 \\ 0 & \frac{2\xi+2\xi^2\psi'(\xi)-1}{1+2\xi} & -\frac{1}{\xi} \\ 0 & -\frac{1}{\xi} & 2\psi'(\xi)-4\psi'(2\xi) \end{bmatrix}. \tag{30}$$

The expression for b can be simplified to

$$b = \phi^2 - 2\phi\kappa\frac{\xi^2}{2\xi+1} + \kappa^2\frac{\xi^3(\xi+1)}{4\xi^2+8\xi+3}, \tag{31}$$

while $a = \phi - \kappa\xi^2/(2\xi+1)$ and $c = 0$.

For small ξ, $b \backsimeq \phi^2 - 4\phi\kappa + (4/3)\kappa^2/\xi$ so the condition $b < 1$ will be violated if ξ is too close to zero when κ is non-zero. On the other hand, letting $\xi \to \infty$ yields

$$b = \phi^2 - 2\phi\kappa + \kappa^2 = (\phi-\kappa)^2$$

and $|\phi - \kappa| < 1$ is the standard invertibility condition for the Gaussian ARMA(1,1) model. Figure 2 shows b plotted against κ for various combinations of parameters.

When ξ is set to 0, the information matrix for $\boldsymbol{\psi}$ and λ is as for a Laplace distribution. On the other hand, if ξ is unknown, the terms in the information matrix that are associated with it go to infinity. A stable information matrix can be obtained by using an exponential link function for ξ. Then, with

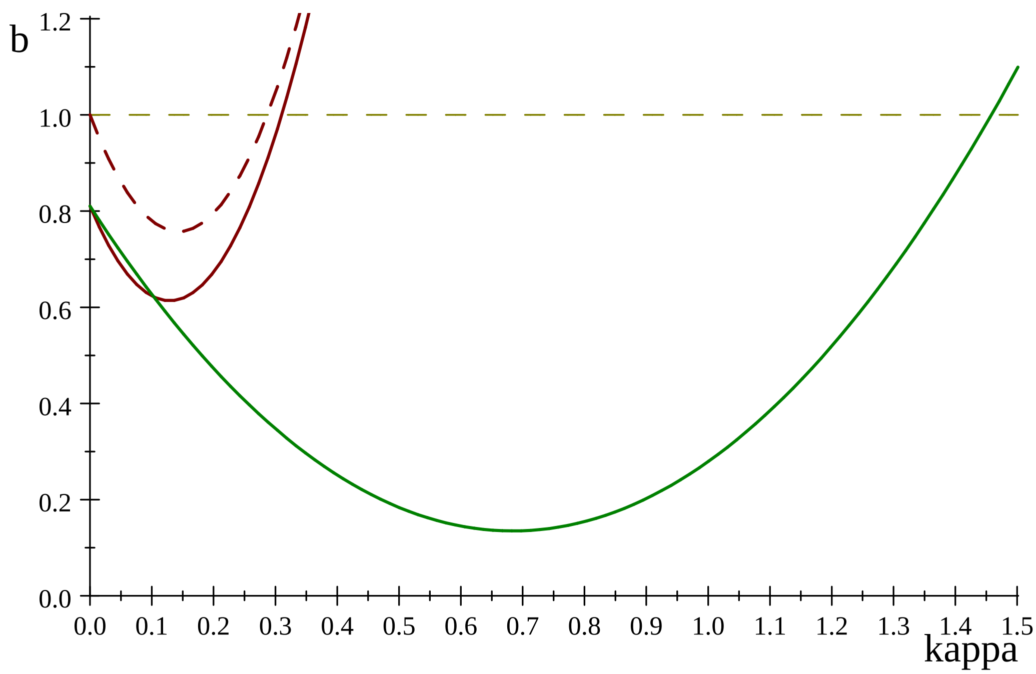


Figure 2: Symmetric EGB2: plot of b against κ for $\phi = 0.9$ with $\xi = 1$ (lower line) and $\xi = 0.1$; dashed line is for $\phi = 1$ with $\xi = 0.1$.

$\xi = \exp(\overline{\xi})$,

$$\mathbf{I}\begin{pmatrix} \boldsymbol{\psi} \\ \lambda \\ \overline{\xi} \end{pmatrix} = \begin{bmatrix} \frac{e^{-2\lambda}\psi'(\xi)\xi^2}{1+2\xi}\mathbf{D}(\boldsymbol{\psi}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{2\xi+2\xi^2\psi'(\xi)-1}{1+2\xi} & -1 \\ \mathbf{0} & -1 & \xi^2[2\psi'(\xi)-4\psi'(2\xi)] \end{bmatrix}, \quad (32)$$

and setting $\xi = 0$ gives

$$\mathbf{I}\begin{pmatrix} \boldsymbol{\psi} \\ \lambda \\ \overline{\xi} \end{pmatrix} = \begin{bmatrix} 2e^{-2\lambda}\mathbf{D}(\boldsymbol{\psi}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1 & -1 \\ \mathbf{0} & -1 & 1 \end{bmatrix}. \quad (33)$$

Unfortunately the matrix is singular, so the problem is not resolved. (This identifiability problem does not arise with the GED shape parameter, υ, perhaps because $\upsilon = 1$ is not on the boundary of the parameter space. But, as noted, the DCS asymptotics do not work for $\upsilon \leq 1.5$.) The implication is that even in a static model, ML estimation may be unstable for shape parameters close to zero. However, small values of these parameters are, in any case, ruled out by the condition $b < 1$.

For the normal distribution matters are more straightforward. Setting $\xi = \overline{\xi}^2$gives

$$\mathbf{I}\begin{pmatrix} \mu \\ \lambda \\ \overline{\xi} \end{pmatrix} = \begin{bmatrix} \frac{2e^{-2\lambda}\psi'(\xi)\xi^2}{1+2\xi} & 0 & 0 \\ 0 & \frac{2\xi+2\xi^2\psi'(\xi)-1}{1+2\xi} & -2\xi^{-1/2} \\ 0 & -2\xi^{-1/2} & 4\xi[\psi'(\xi)-\psi'(2\xi)] \end{bmatrix}, \quad (34)$$

and letting $\xi \to \infty$ yields

$$\mathbf{I}\begin{pmatrix} \mu \\ \lambda \\ \overline{\xi} \end{pmatrix} = \begin{bmatrix} e^{-2\lambda} & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}. \quad (35)$$

Finally note that the hypothesis that $\xi = \varsigma$ is easily tested with a LR statistic. A Wald or Lagrange multiplier (LM) test is also an option. A test of the null hypothesis of Gaussianity can be carried out with a LR test, but with the critical value of the χ_1^2 distribution set at a value corresponding to double the test size; see Chernoff (1954). An LM test may also be carried out as an alternative to the standard Bowman-Shenton test on third and fourth moments. The diagonality of the information matrix in (35) makes the test statistic fairly simple.

5.3 Tests for serial correlation

Before fitting a model, a test against serial correlation may be carried out. Lagrange multiplier (LM) tests against an MA(P) process are based on Ljung-Box (portmanteau) statistics formed using the score. One possibility is to fit an EGB2 distribution to the raw data and to construct scores using the estimated location, scale and shape parameters. An alternative is simply to assume a Laplace distribution, in which case the scores are just indicator variables which take the value one when the observation is above the sample median or minus one if it is below. Thus the information is the same as is used for a runs test and such a test provides an alternative to a test based on sample autocorrelations.

6 Macroeconomic time series

Dynamic location models were fitted to the growth rate of US GDP, industrial production and gross fixed private investment using EGB2, Student's t and normal distributions. GDP and investment data are quarterly, ranging from 1947q1 to 2012q4. Industrial production data are monthly and range from January 1960 to February 2013. All data are seasonally adjusted and taken from the Federal Reserve Economic Data (FRED) database of the Federal Reserve of St. Louis.

When first-order Gaussian models are fitted, there is little indication of residual serial correlation. There is excess kurtosis in all cases, but no evidence of asymmetry. For example, with GDP the Bowman-Shenton statistic is 30.04, which is clearly significant because the distribution under the null hypothesis of Gaussianity is χ^2_2. The non-normality clearly comes from excess kurtosis, which is 1.9, rather than from skewness, which is only 0.18 (with a $p-value$ of 0.24). Comparing the residuals with a fitted normal shows them to have a higher peak at the mean, as well as heavier tails; see Figure 3

The tables below report the estimation results. The Student-t model and the asymmetric EGB2, denoted $EGB2a$, outperform the Gaussian model in all cases with the shape parameters confirming the excess kurtosis and

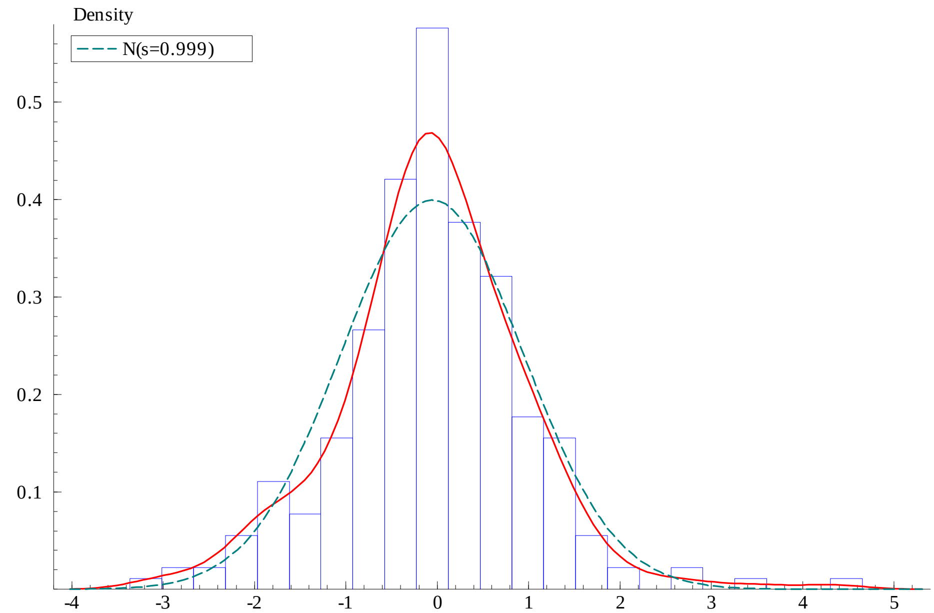


Figure 3: Residuals from fitting a first-order Gaussian model to GDP.

lack of skewness. The EGB2a model is relatively better than the t model with GDP and investment data, according to all the information criteria reported, whereas it is slightly worse with industrial production data. The symmetric EGB2 distribution, called simply $EGB2$, is better in all cases, whether judged by an information criterion or a likelihood ratio test.

The λ parameter is the logarithm of scale, ν, but the estimates of σ are shown because these are comparable across different distributions.

Table 1 **US GDP**

	κ	ϕ	ω	λ	ξ (*or* ν)	ζ	σ	b
EGB2a	0.30	0.50	0.009	−5.40	0.82	0.93		
Num SE	(0.063)	(0.104)	(0.002)	(0.316)	(0.366)	(0.426)	0.0091	0.053
Asy SE	(0.054)	(0.142)	(0.003)	(0.413)	(0.469)	(0.548)		
EGB2	0.30	0.50	0.008	−5.40	0.88			
Num SE	(0.063)	(0.103)	(0.001)	(0.324)	(0.394)	–	0.0091	0.054
Asy SE	(0.054)	(0.143)	(0.001)	(0.415)	(0.502)			
t	0.50	0.50	0.008	−4.88	6.49			
Num SE	(0.094)	(0.103)	(0.001)	(0.071)	(2.364)	–	0.0091	0.049
Asy SE	(0.089)	(0.141)	(0.001)	(0.056)	(1.887)			
Gaussian	0.35	0.49	0.008	−4.70				
Num SE	(0.058)	(0.112)	(0.001)	(0.044)	–	–	0.0091	0.020
Asy SE	(0.061)	(0.141)	(0.001)	(0.044)				

Table 2 **US gross fixed private investment**

	κ	ϕ	ω	λ	ξ (or ν)	ζ	σ	b
EGB2	0.16	0.42	0.019	−4.32	0.35	0.44		
Num SE	(0.055)	(0.156)	(0.008)	(0.445)	(0.170)	(0.252)	0.0525	0.073
Asy SE	(0.045)	(0.234)	(0.008)	(0.514)	(0.206)	(0.271)		
EGB2s	0.14	0.41	0.011	−4.57	0.29			
Num SE	(0.057)	(0.176)	(0.004)	(0.826)	(0.269)	–	0.0525	0.081
Asy SE	(0.043)	(0.261)	(0.004)	(0.614)	(0.200)			
t	0.35	0.38	0.011	−3.24	4.20			
Num SE	(0.114)	(0.171)	(0.004)	(0.085)	(1.275)	–	0.0543	0.049
Asy SE	(0.104)	(0.257)	(0.004)	(0.060)	(0.881)			
Gaussian	0.21	0.36	0.009	−2.95				
Num SE	(0.059)	(0.161)	(0.004)	(0.044)	–	–	0.0525	0.019
Asy SE	(0.062)	(0.257)	(0.004)	(0.044)				

Table 3 **US Industrial production**

	κ	ϕ	ω	λ	ξ (or ν)	ζ	σ	b
EGB2a	0.20	0.86	0.003	−6.07	0.50	0.57		
Num SE	(0.032)	(0.036)	(0.001)	(0.220)	(0.142)	(0.157)	0.0069	0.406
Asy SE	(0.027)	(0.040)	(0.001)	(0.289)	(0.181)	(0.207)		
EGB2	0.20	0.85	0.003	−6.05	0.55			
Num SE	(0.033)	(0.036)	(0.001)	(0.214)	(0.147)	–	0.0069	0.400
Asy SE	(0.027)	(0.040)	(0.001)	(0.287)	(0.196)			
t	0.40	0.85	0.002	−5.25	4.49			
Num SE	(0.060)	(0.036)	(0.001)	(0.046)	(0.743)	–	0.0071	0.396
Asy SE	(0.055)	(0.040)	(0.001)	(0.038)	(0.634)			
Gaussian	0.25	0.83	0.002	−4.95				
Num SE	(0.032)	(0.041)	(0.001)	(0.028)	–	–	0.0071	0.343
Asy SE	(0.035)	(0.046)	(0.001)	(0.028)				

Table 4 **US macroeconomic series - Model comparison**

	Log-Likelihood	**AIC**	**BIC**
	GDP		
EGB2a	868.548	-6.559	-6.478
EGB2	868.376	-6.566	-6.498
t	868.242	-6.565	-6.497
Gaussian	862.212	-6.526	-6.472
	Gross fixed private investment		
EGB2a	414.032	-3.103	-3.021
EGB2	416.001	-3.125	-3.058
t	411.91	-3.094	-3.026
Gaussian	401.947	-3.026	-2.972
	Industrial production		
EGB2a	2292.18	-7.167	-7.125
EGB2	2291.66	-7.168	-7.133
t	2293.56	-7.174	-7.139
Gaussian	2255.21	-7.057	-7.029

Figure 4 shows the investment series with the filtered estimates of location, $\mu_{t|t-1}$, for the symmetric EGB2, Gaussian and t-models. As can be seen the main difference is in the first quarter of 2009, where the response from the Gaussian model is more pronounced.

7 Structural breaks

It might be thought that the EGB2 and t filters will be less responsive to a permanent change in the level than the linear Gaussian filter. However, for moderate size shifts, the score functions in Figure 1 suggest that this might not be the case, because only for large observations is the Gaussian response bigger than the response of the robust filters. For example, for the logistic (EGB2 with unit shape parameters), the score is only smaller than the observation (and hence the linear filter) when it is more than (approximately) 1.6 standard deviations from the mean. The behaviour of the t-filter is similar.

In order to investigate the issue of adapting to a permanent change in level, an upward shift was added to the US industrial production data at the begining of 2010. The size of the shift was calibrated so as to be proportional

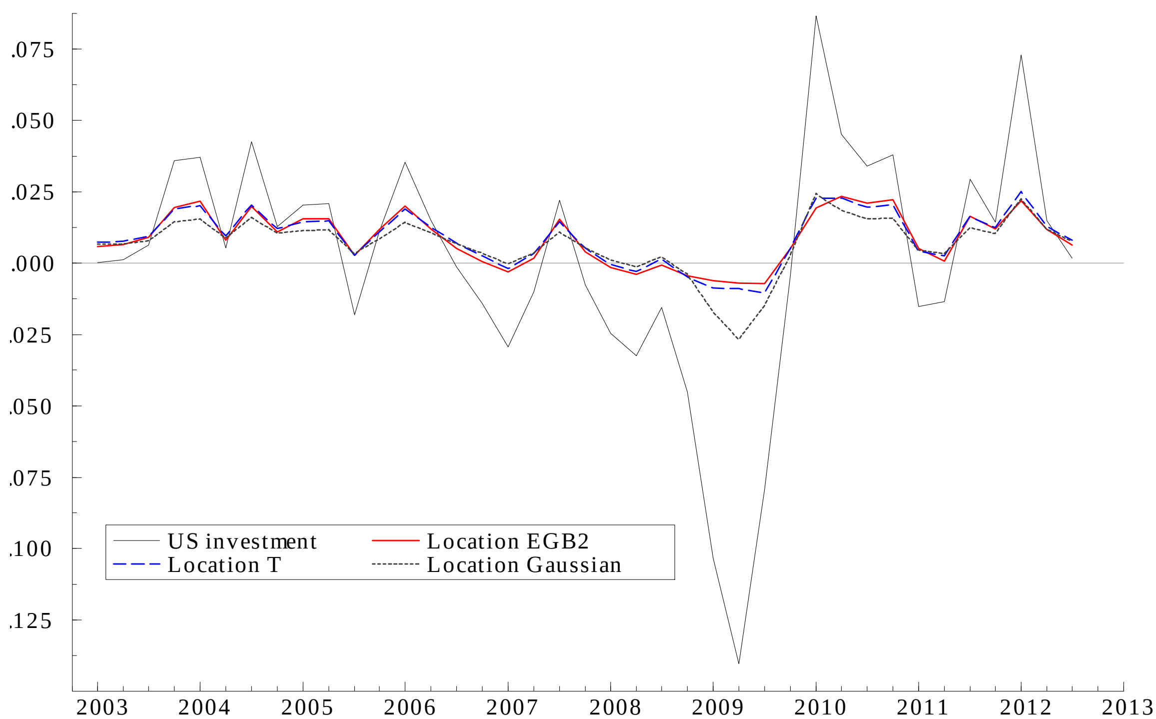


Figure 4: Estimates of location for US Investment

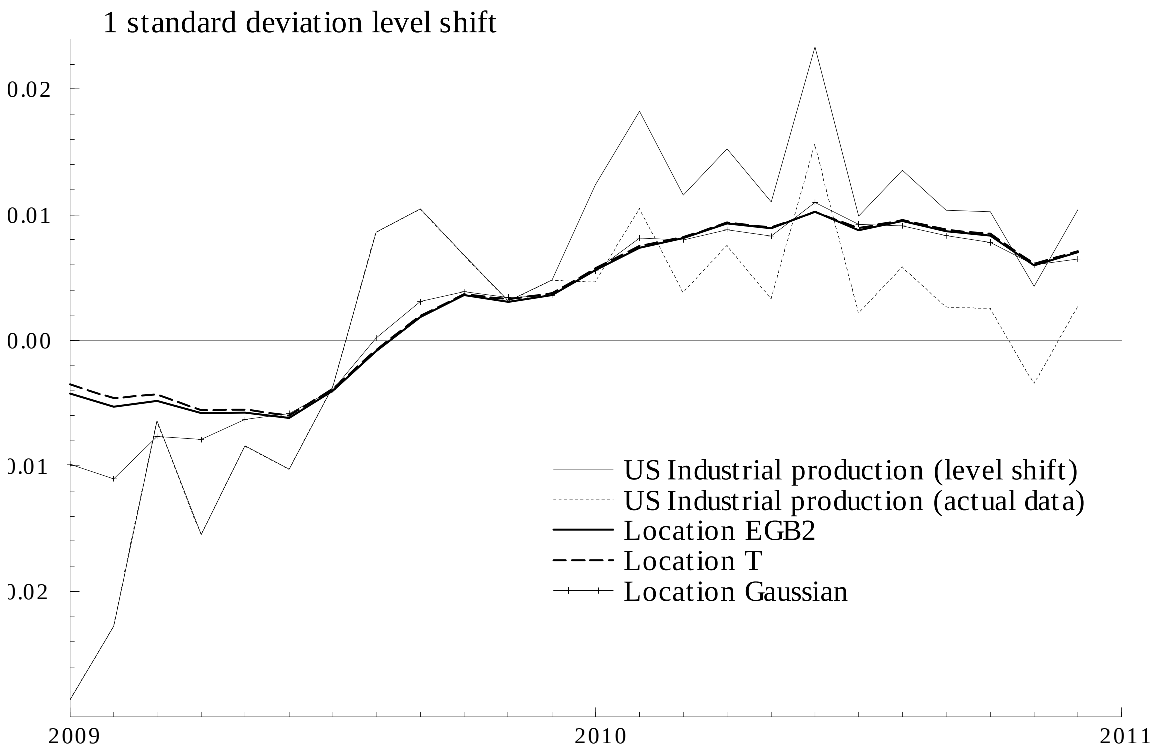


Figure 5: Response of Gaussian, EGB2 and t-filters to a one SD shift in level in January 2010

to the sample standard deviation of the series. The results are shown in the Figures 5 and 6. For a one standard deviation shift the paths of all three filters after the break are very similar. Only for two standard deviations does the Gaussian filter adapt more quickly.

8 EGB2-EGARCH

Dynamic scale models can be constructed for conditional t and GED distributions; see Harvey (2013, ch 4). In the former case the score has a beta distribution, whereas in the latter, it has a gamma distribution. Just as in the dynamic location case, the EGB2 distribution offers an alternative to the GED for capturing responses between the normal and Laplace. Wang et al

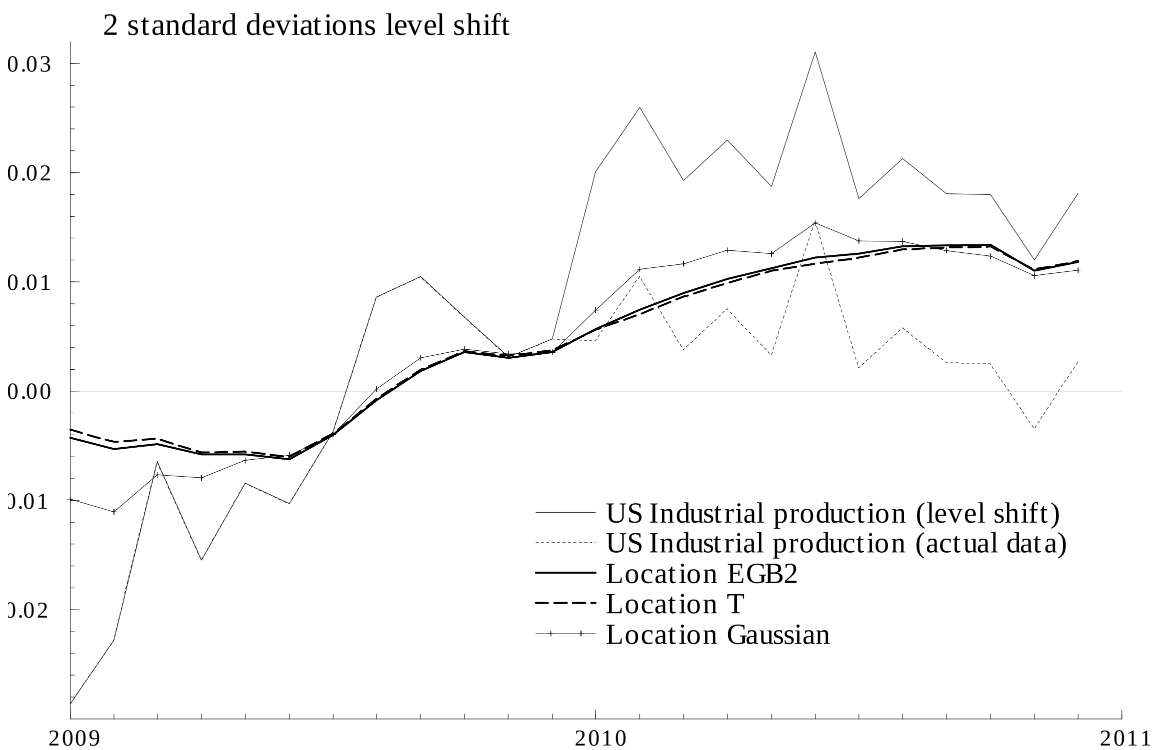


Figure 6: Response of Gaussian, EGB2 and t-filters to a two SD shift in level in January 2010

(2001) fitted GARCH-EGB2 models to daily US dollar exchange rate data for six major currencies and found evidence to favor them over the conventional GARCH-t and EGARCH-t alternatives, particularly for exchange rate data characterized by skewness. The EGARCH-EGB2 model therefore needs to be taken seriously: its theoretical properties are set out below.

The first-order dynamic scale model with EGB2 distributed errors is

$$y_t = \mu + \exp(\lambda_{t|t-1})\varepsilon_t, \qquad t = 1, ..., T, \tag{36}$$

where ε_t is a standardized ($\mu = 0, \nu = 1$) EGB2, that is $\varepsilon_t \sim EGB2(0, 1, \xi, \varsigma)$. The dynamic equation is

$$\lambda_{t+1|t} = \omega(1 - \phi) + \phi\lambda_{t|t-1} + \kappa u_t, \tag{37}$$

where u_t is now the score with respect to $\lambda_{t|t-1}$. The conditional distribution is

$$f_t(y_t \mid Y_{t-1}; \mu, \boldsymbol{\psi}, \xi, \varsigma) = \frac{\exp\{\xi(y_t - \mu)e^{-\lambda_{t|t-1}}\}}{e^{\lambda_{t|t-1}}B(\xi, \varsigma)(1 + \exp\{(y - \mu)e^{-\lambda_{t|t-1}}\})^{\xi+\varsigma}},$$

where $\boldsymbol{\psi}$ now denotes the parameters in (37). The conditional score is

$$u_t = \frac{\partial \ln f(y_t)}{\partial \lambda_{t|t-1}} = (\xi + \varsigma)\varepsilon_t b_t - \xi\varepsilon_t - 1, \tag{38}$$

and

$$b_t = \frac{\exp\{(y - \mu)e^{-\lambda_{t|t-1}}\}}{1 + \exp\{(y - \mu)e^{-\lambda_{t|t-1}}\}} = \frac{\exp \varepsilon_t}{1 + \exp \varepsilon_t}.$$

At the true parameters values, $b_t \sim beta(\xi, \varsigma)$ as in the score for the dynamic location model.

The model may be parameterized in terms of the standard deviation, $\sigma_{t|t-1}$, by defining $\epsilon_t = \varepsilon_t/h$. Then

$$y_t = \mu + \exp(\lambda_{\sigma,t|t-1})\epsilon_t, \qquad t = 1, ..., T,$$

with the only difference between $\lambda_{\sigma,t|t-1}$ and $\lambda_{t|t-1}$ being in the constant term which in $\lambda_{\sigma,t|t-1}$ is $\omega_\sigma = \omega + \ln h$; see the earlier discussion in sub-section 5.1. Note that the variance of ϵ_t is unity.

Writing the score, (38) as

$$u_t = h(\xi + \varsigma)\epsilon_t b_t - h\xi\epsilon_t - 1, \tag{39}$$

it can be seen[8] that when $\xi = \varsigma = 0$, $u_t = \sqrt{2}\,|\epsilon_t| - 1$ and, when $\xi = \varsigma \to \infty$, $u_t = \epsilon_t^2 - 1$.

Figure 7 compares the way observations are weighted by the score of a EGB2 distribution with $\xi = \varsigma = 0.5$, a Student's t_7 distribution and a $GED(1.148)$. These are the same distributions used in the earlier figures; all have excess kurtosis of 2. Note that the relationship of the above EGB2 scale score to the location score is as implied by (11). Consistent with this relationship and the Winsorizing of the location score, dividing (39) by ϵ_t gives a bounded function as $|\epsilon_t| \to \infty$.

The following result, which is related to Lemma 1 of Harvey (2013, p23), is useful for deriving the asymptotic properties of the maximum likelihood estimator.

Proposition 3 *If $\epsilon_t \sim EGB2(0, 1, \xi, \varsigma)$, then for h and $k \geq 0$,*

$$E_{EGB2(0,1,\xi,\varsigma)}[\varepsilon_t^r b_t^h (1-b_t)^k] = \frac{B(\xi + h, \varsigma + k)}{B(\xi, \varsigma)} E_{EGB2(0,1,\xi+h,\varsigma+k)}[\varepsilon_t^r], \quad r = 1, 2, 3, ... \tag{40}$$

Proof. The result follows from writing

$$\begin{aligned} E_{EGB2(0,1,\xi,\varsigma)}[\varepsilon_t^r b_t^h (1-b_t)^k] &= \int \frac{\nu\varepsilon_t^r}{B(\xi,\varsigma)} \frac{\exp h\varepsilon_t}{(1+\exp\varepsilon_t)^{h+k}} \frac{\exp\xi\varepsilon_t}{(1+\exp\varepsilon_t)^{\xi+\varsigma}} d\varepsilon_t \\ &= \frac{B(\xi + h, \varsigma + k)}{B(\xi,\varsigma)} \int \frac{\nu\varepsilon_t^r}{B(\xi + h, \varsigma + k)} \frac{\exp(\xi + h)\varepsilon_t}{(1+\exp\varepsilon_t)^{\xi+\varsigma+h+k}} d\varepsilon_t \end{aligned}$$

■

[8] When $\xi = 0$, $\xi h = \sqrt{2}$ and b_t degenerates to a Bernoulli variable such that $b_t = 0$ when $\epsilon_t < 0$ and $b_t = 1$ when $\epsilon_t > 0$. Then $2b_t - 1 = 1$ (-1) for $\epsilon_t > 0$ $(\epsilon_t < 0)$ and the score can be written as: $u_t = \sqrt{2}\,|\epsilon_t| - 1$.

As regards $\xi \to \infty$, note that because $\partial b_t/\partial\epsilon_t = hb_t(1-b_t)$, a first order Taylor expansion of b_t around $\varepsilon_t = 0$ yields $b_t \simeq \frac{1}{2} + \frac{h}{4}\epsilon_t$.Therefore $2b_t - 1 \simeq (h/2)\epsilon_t$ and $u_t \simeq (\xi h^2/2)\epsilon_t^2 - 1$. As $\xi \to \infty$, $\xi h^2 \to 2$.

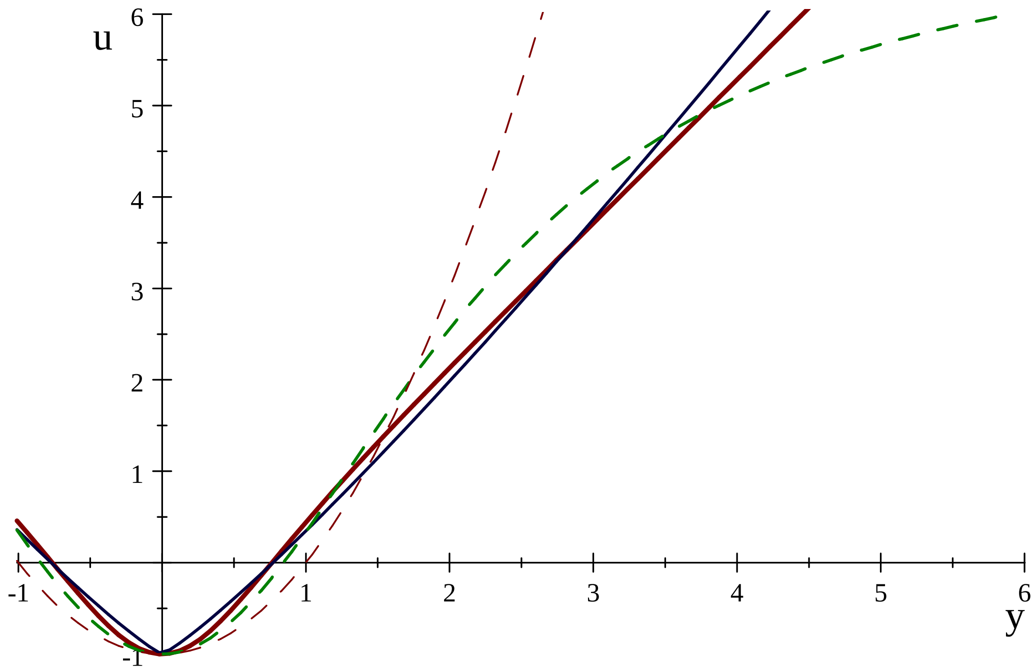


Figure 7: Score functions for EGB2 (thick line), GED (medium line) and t (thick dash), all with excess kurtosis of 2. Thin line shows normal score.

The above result can be used to confirm that $E(u_t) = 0$. (Note that $\psi(\xi+1) = \psi(\xi) + 1/\xi$).

The unconditional mean is given by $E(y_t) = \mu + E(\varepsilon_t) E(e^{\lambda_{t|t-1}})$, whereas the $m-th$ unconditional moment about the mean is $E(\varepsilon_t^m) E(e^{m\lambda_{t|t-1}})$, $m > 1$. In the Beta-t-EGARCH and Gamma-GED-EGARCH models analysed in Harvey (2013, ch. 4), the expression $E(\exp(m\lambda_{t|t-1}))$ depends on the moment generating function (MGF) of a beta variate which has a known form. For EGB2-EGARCH, the unconditional moments depend on the MGF of u_t, ie $E_{EGB2(\xi,\varsigma)}[mu_t]$, where u_t is defined in (38). For the limiting normal and Laplace cases of the EGB2, the score functions and hence the unconditional moments are the same as for $\upsilon = 2$ and $\upsilon = 1$ in Gamma-GED-EGARCH; see Harvey (2013, sub-section 4.2.2). For $\upsilon = 1$ it is necessary to have $m\kappa < 1$ in the first-order model for the $m-th$ moment to exist, whereas for $\upsilon = 2$ the condition is $m\kappa < 1/2$. For $0 < \xi, \varsigma < \infty$ having the last condition hold is therefore sufficient for the existence of the unconditional moments. This being the case, we can at least assert, using Jensen's inequality, that the unconditional moments exceed the conditional moments and that the kurtosis increases; see Harvey (2013, p 102).

The MGF of u_t is also required to find the conditional expectations needed to forecast volatility and volatility of volatility. However, it is the full $\ell - step$ ahead conditional distribution that is often needed in practice and this is easily simulated from standardized beta variates. The quantiles, such as those needed for VaR and the associated expected shortfalls, may be estimated at the same time.

8.1 Maximum likelihood estimation

The asymptotic distribution of the ML estimators of the parameters in a dynamic scale model with an EGB2 distribution is given in the proposition below. In the general asymmetric case, it is assumed that μ is given, because the cross-terms of the information matrix associated with it and the other parameters depend on scale.

Because both the score and its derivatives are just linear combinations of variables of the form $\varepsilon_t^r b_t^h (1-b_t)^k$ it is not difficult to show that the conditions for convergence and asymptotic normality of the maximum likelihood estimator are satisfied.

Proposition 4 *Consider model defined by (36) and (37) with $|\phi| < 1$ and μ assumed to be known. Define a, b and c as in (14) with*

$$
\begin{aligned}
E(u_t') &= -\left(\frac{1}{\xi+\varsigma+1}\left(\psi'(\xi)+\psi'(\varsigma)+\left(\psi(\xi)-\psi(\varsigma)-\frac{\xi-\varsigma}{\xi\varsigma}\right)^2-\frac{\xi^2+\varsigma^2}{\xi^2\varsigma^2}\right)+1\right) = -\sigma_u^2 \\
E(u_t'^2) &= \frac{\xi\varsigma\left(\xi+1\right)\left(\varsigma+1\right)\left(\xi+\varsigma\right)}{\left(\xi+\varsigma+3\right)\left(\xi+\varsigma+2\right)\left(\xi+\varsigma+1\right)}\boldsymbol{\gamma}_1\left(\xi+2,\varsigma+2\right) \\
&\quad +\frac{2\xi\varsigma\left(\xi+1\right)\left(\xi+\varsigma\right)}{\left(\xi+\varsigma+2\right)\left(\xi+\varsigma+1\right)}\boldsymbol{\gamma}_2\left(\xi+2,\varsigma+1\right)-\frac{2\xi^2\varsigma}{\xi+\varsigma+1}\boldsymbol{\gamma}_2\left(\xi+1,\varsigma+1\right)+\sigma_u^2+1 \\
E(u_t u_t') &= -\frac{\xi\varsigma\left(\xi+1\right)\left(\xi+\varsigma\right)}{\left(\xi+\varsigma+2\right)\left(\xi+\varsigma+1\right)}\boldsymbol{\gamma}_2\left(\xi+2,\varsigma+1\right)+\frac{\xi^2\varsigma}{\xi+\varsigma+1}\boldsymbol{\gamma}_2\left(\xi+1,\varsigma+1\right)-1,
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{\gamma}_1\left(p,q\right) &= \psi'''(p)+\psi'''(q)+3\left(\psi'(p)+\psi'(q)\right)^2+4\left(\psi''(p)-\psi''(q)\right)\left(\psi(p)-\psi(q)\right) \\
&\quad +6\left(\psi'(p)+\psi'(q)\right)\left(\psi(p)-\psi(q)\right)^2+\left(\psi(p)-\psi(q)\right)^4, \\
\boldsymbol{\gamma}_2\left(p,q\right) &= \psi''(p)-\psi''(q)+3\left(\psi(p)-\psi(q)\right)\left(\psi'(p)+\psi'(q)\right)+\left(\psi(p)-\psi(q)\right)^3
\end{aligned}
$$

Let $\boldsymbol{\psi} = (\kappa,\ \phi,\ \omega)'$, where $\omega = \omega_\sigma - \ln h$. Assuming that $b < 1$ and $\kappa \neq 0$, $(\widetilde{\boldsymbol{\psi}}', \widetilde{\xi}, \widetilde{\varsigma})'$, the ML estimator of $(\boldsymbol{\psi}', \xi, \varsigma)'$, is consistent and the limiting distribution of $\sqrt{T}((\widetilde{\boldsymbol{\psi}} - \boldsymbol{\psi})', \widetilde{\xi} - \xi, \widetilde{\varsigma} - \varsigma)'$ is multivariate normal with mean vector zero and covariance matrix given by $Var(\widetilde{\boldsymbol{\psi}}, \widetilde{\xi}, \widetilde{\varsigma}) = \mathbf{I}^{-1}(\boldsymbol{\psi}, \xi, \varsigma)$, where the information matrix is

$$
\mathbf{I}\begin{pmatrix}\boldsymbol{\psi}\\ \xi \\ \varsigma\end{pmatrix} = \begin{bmatrix} I_{22}\mathbf{D}(\boldsymbol{\psi}) & I_{23}\mathbf{d} & I_{24}\mathbf{d} \\ I_{23}\mathbf{d}' & \psi'(\xi)-\psi'(\xi+\varsigma) & -\psi'(\xi+\varsigma) \\ I_{24}\mathbf{d}' & -\psi'(\xi+\varsigma) & \psi'(\varsigma)-\psi'(\xi+\varsigma) \end{bmatrix}
$$

with I_{22}, I_{23} and I_{24} as in Proposition 2 and $\mathbf{D}(\boldsymbol{\psi})$ and $\mathbf{d}$ are defined as in (17) with $k = 1$.

Proof. The first derivative with respect to $\lambda_{t|t-1}$ is:

$$
u_t' = -(\xi+\varsigma)[\varepsilon_t^2 b_t(1-b_t)+\varepsilon_t b_t]+\xi\varepsilon_t
$$

Proposition 3 can be used to evaluate

$$E(u_t') = -(\xi + \varsigma)[E(\varepsilon_t^2 b_t(1 - b_t)) - E(\varepsilon_t b_t) + \frac{\xi}{\xi + \varsigma} E(\varepsilon_t)]$$

$$\begin{aligned} E(u_t'^2) &= (\xi + \varsigma)^2 [E(\varepsilon_t^2 b_t^2) + E(\varepsilon_t^4 b_t^2 (1 - b_t)^2) + 2E(\varepsilon_t^3 b_t^2 (1 - b_t))] \\ & \quad + \xi^2 E(\varepsilon_t^2) - 2\xi(\xi + \varsigma)[E(\varepsilon_t^3 b_t (1 - b_t)) + E(\varepsilon_t^2 b_t)] \end{aligned}$$

and $E(u_t' u_t)$.

The second and third derivatives with respect to $\lambda_{t|t-1}$ are:

$$u_t'' = (\xi + \varsigma)[3\varepsilon_t^2 b_t(1 - b_t) + \varepsilon_t^3 b_t (1 - b_t)^2 - \varepsilon_t^3 b_t^2 (1 - b_t) + \varepsilon_t b_t] - \xi \varepsilon_t$$

and

$$\begin{aligned} u_t''' &= -(\xi + \varsigma)[7\varepsilon_t^2 b_t (1 - b_t) + 3\varepsilon_t^3 b_t (1 - b_t)^2 - 3\varepsilon_t^3 b_t^2 (1 - b_t) + 12\varepsilon_t^4 b_t^2 (1 - b_t)^2 \\ & \quad - 3\varepsilon_t^4 b_t (1 - b_t)^3 - 3\varepsilon_t^4 b_t^3 (1 - b_t) + \varepsilon_t b_t] + \xi \varepsilon_t \end{aligned}$$

As can be seen, the first three derivatives of the score are linear combinations of variables $\varepsilon_t^r b_t^h (1 - b_t)^k$, with $r, h, k \in \{0, 1, 2, 3, 4\}$. Therefore, in the light of Proposition 3 , their expected values exist and are time invariant. Furthermore, as any power of such derivatives and their cross-products is also a linear combination of $\varepsilon_t^r b_t^h (1 - b_t)^k$ variables, all moments exist and are time invariant.

Consistency and asymptotic normality can be proved by showing that the conditions for Lemma 1 in Jensen and Rahbek (2004, p 1206) hold. The final condition, the boundedness of the third derivative in the region of the true parameter values, follows partly from an argument similar to the one employed for the DCS location-t model; see Harvey (2013, p 66). The key point here is that terms of the form $\varepsilon_t^r b_t^h (1 - b_t)^k$, h or k greater than zero, are bounded for all admissible parameter values because $\varepsilon_t^r b_t^h (1 - b_t)^k \to 0$ as $y_t \to \pm\infty$. The final term, $\xi \epsilon_t$, can be bounded from above[9]; see Harvey (2013, p 44-5). ■

[9] Because all moments of the random variable $\epsilon_t(\boldsymbol{\psi}_0)$ exist, all moments of $\epsilon_t(\boldsymbol{\psi})$ exist in the neighborhood of $\boldsymbol{\psi}_0$ because $\epsilon_t(\boldsymbol{\psi})$ is equal to $\epsilon_t(\boldsymbol{\psi}_0)$ times a scalar that is independent of y_t and is given by $\exp \lambda_{t|t-1}(\boldsymbol{\psi}_0 - \boldsymbol{\psi})$, where the notation $\lambda_{t|t-1}(\boldsymbol{\psi}_0 - \boldsymbol{\psi})$ indicates that $\lambda_{t|t-1}$ depends on $\boldsymbol{\psi}_0 - \boldsymbol{\psi}$. If a value $\boldsymbol{\psi}^*$ is chosen so that $\lambda_{t|t-1}(\boldsymbol{\psi}_0 - \boldsymbol{\psi}^*) > \lambda_{t|t-1}(\boldsymbol{\psi}_0 - \boldsymbol{\psi})$, for all $t = 1, ..., T$, then $\epsilon_t(\boldsymbol{\psi}^*)$ will uniformly bound $\epsilon_t(\boldsymbol{\psi})$ from above.

Remark 2 *Note that $E(u_t') = -I_{22}$ in Proposition 2, as it should be.*

Remark 3 *As with the Beta-t-EGARCH model, the asymptotic theory adapts straightforwardly when ϕ is set equal to one; see Harvey (2013, p 117).*

8.2 Symmetric distribution

The properties of the ML estimator simplify considerably in the symmetric case as the following corollary to Proposition 4 shows.

Corollary 1 *When it is known that $\xi = \varsigma$, the expressions needed to obtain a, b and c simplify to:*

$$E(u_t') = \frac{1 - 2\xi^2\psi'(\xi) - 2\xi}{2\xi + 1} = -\sigma_u^2 \tag{41}$$

$$E(u_t'^2) = \frac{\xi^3(\xi+1)}{(2\xi+3)(2\xi+1)}(2\psi'''(\xi+2) + 12\psi'^2(\xi+2)) + \sigma_u^2 + 1 \tag{42}$$

and

$$E(u_t u_t') = -1. \tag{43}$$

When $\xi = 0$, so that the distribution is Laplace, $E(u_t') = -1$. Similarly as $\xi \to \infty$, $E(u_t') = -2$, which is the correct result for a Gaussian distribution. In addition, when $\xi = 0$ both $\psi'(\xi+2)$ and $\psi'''(\xi+2)$ are finite so $E(u_t'^2) = 2$. Hence

$$b = \phi^2 - 2\phi\kappa + 2\kappa^2, \tag{44}$$

which is the same as given by the expression in Harvey (2013, p 120) for b in Gamma-GED-EGARCH when $\upsilon = 1$. (Also $c = -1$.) Similarly for $\xi \to \infty$,

$$b = \phi^2 - 4\phi\kappa + 12\kappa^2.$$

Remark 4 *As can be seen from (30), the information matrix is block diagonal in the symmetric model, so μ can be included in the set of parameters to be estimated by ML without affecting the asymptotic distribution of $(\widetilde{\boldsymbol{\psi}}', \widetilde{\xi})$.*

8.3 Exchange rate application

Beta-t-EGARCH and EGB2-EGARCH models were fitted to daily exchange rate returns, defined as the first difference of the logarithm of the exchange rate. for the Swiss Franc against the Euro for 4th January 1999 to 15th March 2013. The results are shown in Table 5. The t-distribution gives a better fit. The graph in Figure 8 shows a subset of the absolute values of the returns together with the filtered estimates of the standard deviation adjusted so as to coincide with the observations rather than being one-step ahead. The observation for period 104 has been truncated because its value is 0.09. As can be seen from the graph, the filter for the Beta-t-EGARCH model responds somewhat less to this observation than does the EGB2-EGARCH filter. The response of the standard GARCH-t model is much bigger and it takes over thirty periods to come back to the level of the EGARCH filters.

In this particular instance, the outlier has a clear explanation. On 6th September 2011, the Swiss National Bank announced its intention to enforce a ceiling on the exchange rate of the euro against the Swiss franc. As a result, the exchange rate experienced a sudden fall. After removing the resulting outlier from the returns series, the EGB2 performs better than the Student's t. For both data sets GARCH-t has a much smaller log-likelihood.

Table 5 **ML estimates for the Swiss franc**

A: 4/1/1999 to 15/3/2013 **B:** 6/9/2011 removed

	EGB2					Student's t					GARCH-t
	κ	ϕ	ω	ξ	Log-L	κ	ϕ	ω	ν	Log-L	Log-L
A	0.018	0.993	-5.56	1.05	12848.2	0.017	0.994	-5.14	8.47	12849.9	12838.7
B	0.017	0.994	-5.47	1.22	12865.0	0.016	0.994	-5.13	9.69	12863.1	12856.6

8.4 Modeling returns with the martingale difference modification

There is a problem with modeling returns with a skewed distribution because the conditional expectation,

$$E_{t-1}y_t = \mu_\varepsilon \exp(\lambda_{t|t-1}),$$

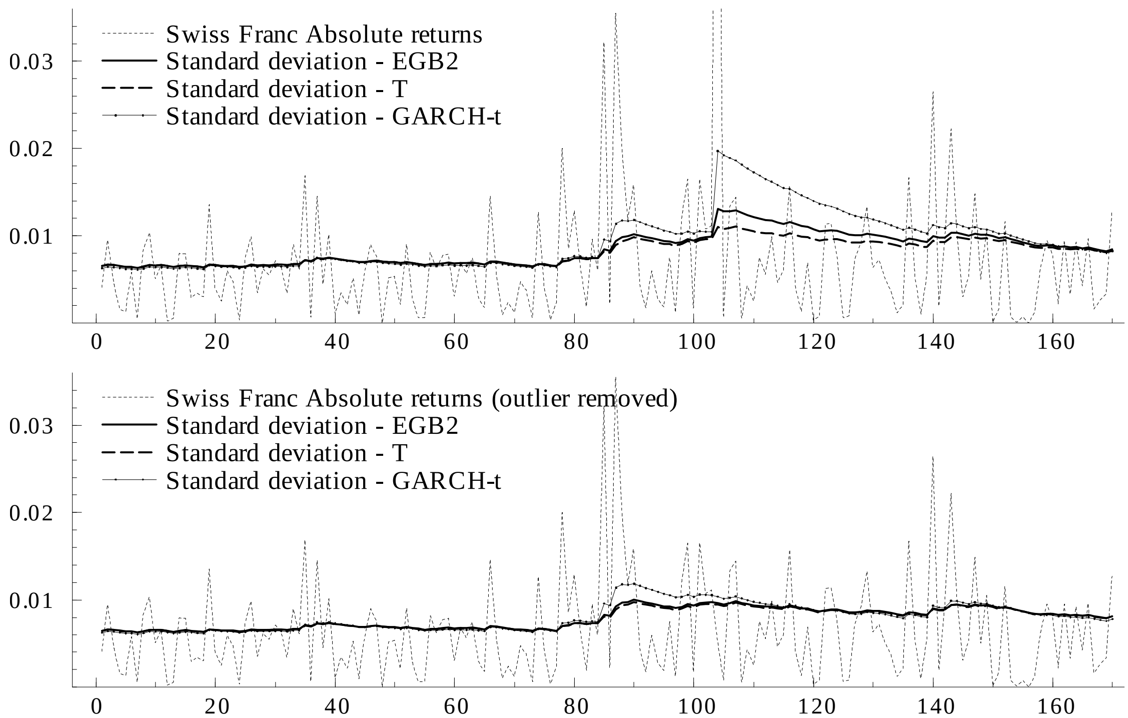


Figure 8: Swiss franc returns against the Euro with EGARCH and GARCH filters.

is not constant. Therefore y_t cannot be a martingale difference. Following Harvey and Sucarrat (2012), the model is re-formulated as

$$y_t = (\varepsilon_t - \mu_\varepsilon) \exp(\lambda_{t|t-1}), \quad t = 1, ..., T, \tag{45}$$

where $\mu_\varepsilon = \psi(\xi) - \psi(\varsigma)$; see also Giot and Laurent (2003) who transform their Skew-t GARCH model to make it a MD.

The score is now

$$u_t = (\xi + \varsigma) b_t(\varepsilon_t - \mu_\epsilon) - \xi(\varepsilon_t - \mu_\epsilon) - 1, \tag{46}$$

Then $E(\varepsilon_t - \mu_\varepsilon) = 0$, but[10] $E(b_t(\varepsilon_t - \mu_\varepsilon)) = (\xi/(\xi + \varsigma)) E_{EGB2(\xi+1,\varsigma)}[\varepsilon_t] - (\xi/(\xi+\varsigma))\mu_\varepsilon = 1$. Hence $E(u_t) = 0$.

Although any analytic derivations are more complicated than in Proposition 4, the important point to note is that the distribution of u_t does not depend on $\lambda_{t|t-1}$ and neither do the distributions of its derivatives. Thus the conditions for the ML estimator to be consistent and asymptotically normal still hold.

9 Changing location and changing scale

The DCS model for time-varying location may be combined with a DCS EGARCH model to give

$$y_t = \mu_{t|t-1} + \exp(\lambda_{t|t-1})\varepsilon_t, \quad t = 1, ..., T.$$

For symmetric distributions, the structure of the information matrix in the static model is such that the form of the dynamic equations for $\mu_{t|t-1}$ and $\lambda_{t|t-1}$ is essentially unchanged, except that both scores now contain $\lambda_{t|t-1}$ and $\mu_{t|t-1}$. Estimation by ML is straightforward. Unfortunately the presence of $\lambda_{t|t-1}$ in the part of the information matrix associated with $\mu_{t|t-1}$ means that the asymptotic theory developed when either location or scale is dynamic cannot be applied[11]. Some other route is needed to establish consistency and asymptotically normality of the ML estimators.

The following tables report results for the quarterly rate of CPI inflation in

[10] Or $E_{EGB2(-\mu_\varepsilon,1,\xi+1,\varsigma)}[\varepsilon_t - \mu_\varepsilon]$

[11] Asymptotic results are similarly difficult to prove for the ARMA-GARCH model; see Francq and Zakoian (2010) and Lange et al (2010).

the UK (expressed in annualized percentage terms). The data are taken from the Federal Reserve Economic Data (FRED) database of the Federal Reserve of St. Louis and seasonally adjusted using trigonometric seasonals with the STAMP package of Koopman, Harvey, Doornik and Shephard (2009). The sample ranges from 1955q1 to 2013q1 and the estimation has been carried out both on the full sample and on a sample restricted to the post-1983 years, which roughly coincide with a period of less pronounced inflation variability (the 'Great Moderation'). A local level model is used for location, that is

$$\mu_{t+1|t} = \mu_{t|t-1} + \kappa^{\dagger} u_{\mu,t}, \quad t = 1, ..., T,$$

where $\kappa^{\dagger}$ is used simply to differentiate it from the corresponding coefficient for scale.

A dynamic location model was first fitted, giving the results in Table 6. Tests based on the first-order autocorrelations of squared residuals show that the null hypothesis of homoscedasticity is convincingly rejected. (The χ_1^2 statistics are 11.60, 11.78 and 12.12 for residuals from EGB2, t and Gaussian models respectively).

Table 6 **UK quarterly CPI inflation - sample: 1955q1-2013q1**

	$\kappa^{\dagger}$	λ	ξ (or ν)	ς	$Log-L$	BIC
EGB2a	0.34 (0.041)	−0.24 (0.34)	0.57 (0.262)	0.33 (0.130)	−565.17	4.97
EGB2	0.46 (0.073)	0.25 (0.248)	0.73 (0.245)	–	−574.2	5.02
t	0.83 (0.161)	0.71 (0.070)	3.14 (0.567)	–	−571.3	5.00
Gaussian	0.49 (0.070)	1.12 (0.046)	–	–	−589.5	5.13

Tables 7 and 8 show results for the full model with time-varying location and scale. Numerical standard errors are given in parenthesis. Over the full sample UK inflation displays several extreme observations, and estimates based on distributions allowing for excess kurtosis perform better, with the EGB2 outperforming both the Gaussian and the Student t model. A plot of the dynamic levels[12] over the 1970s shows that, contrary to EGB2 and t, the Gaussian estimate reacts strongly to extreme observations; see Figure 9.

A histogram of inflation residuals from the EGB2 model reveals non-

[12] The plot shows the contemporaneous filter, ie $\mu_{t+1|t}$ plotted against t.

negligible asymmetry; see Figure 10. The asymmetry can be successfully accomodated by the unrestricted EGB2 model, with the null hypothesis of symmetry being clearly rejected by a likelihood ratio test. In the period after 1982, when inflation is much lower, both asymmetry and excess kurtosis are less pronounced; even the Gaussian model performs reasonably well, but the EGB2 still gives the best fit.

Table 7 **UK quarterly CPI inflation - sample: 1955q1-2013q1**

	$\kappa^{\dagger}$	κ	ϕ	δ	ξ (or ν)	ς	$Log-L$	BIC
EGB2a	0.24 (0.046)	0.09 (0.026)	0.99 (0.014)	-0.004 (0.008)	0.99 (0.474)	0.48 (0.170)	-526.8	4.68
EGB2	0.27 (0.054)	0.11 (0.029)	0.99 (0.016)	-0.01 (0.01)	0.51 (0.161)	–	-532.8	4.71
t	0.55 (0.109)	0.13 (0.036)	0.98 (0.024)	0.01 (0.02)	4.58 (1.223)	–	-536.2	4.74
Gaussian	0.39 (0.070)	0.05 (0.010)	0.96 (0.015)	0.03 (0.02)	–	–	-568.3	4.99

Table 8 **UK quarterly CPI inflation - sample: 1983q1-2013q1**

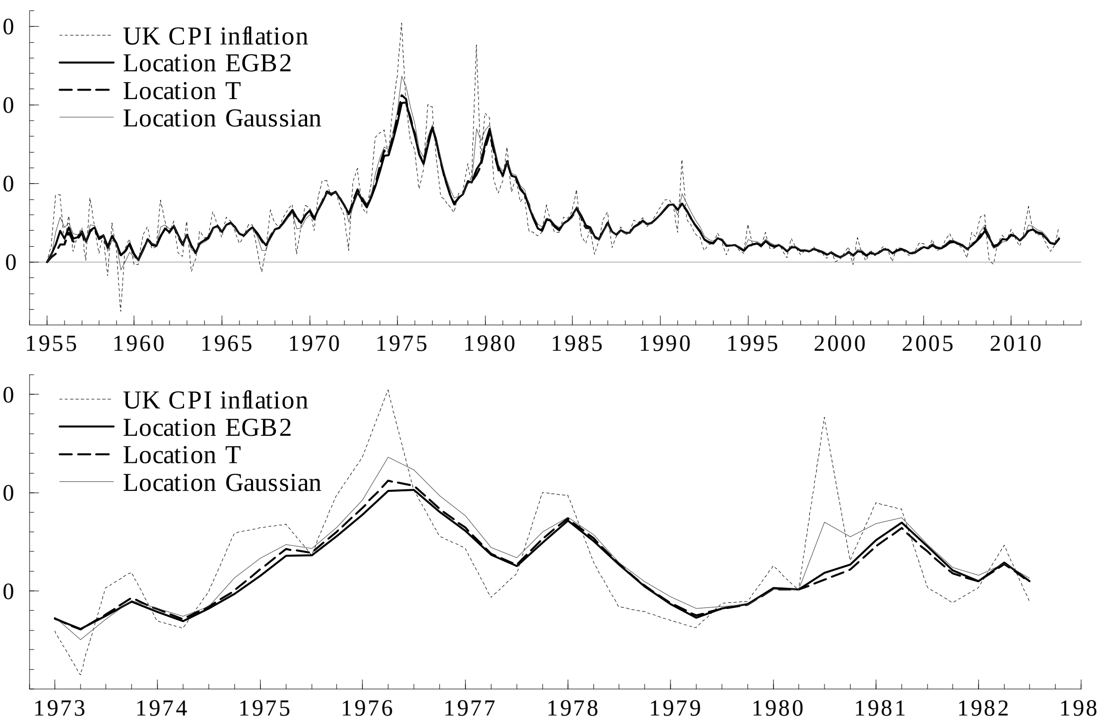
	$\kappa^{\dagger}$	κ	ϕ	δ	ξ (or ν)	ς	$Log-L$	BIC
EGB2a	0.19 (0.051)	0.14 (0.079)	0.91 (0.075)	-0.14 (0.14)	0.33 (0.272)	0.19 (0.139)	-216.9	3.85
EGB2	0.21 (0.106)	0.19 (0.083)	0.91 (0.064)	-0.15 (0.12)	0.20 (0.15)	–	-220.5	3.94
t	0.48 (0.141)	0.20 (0.058)	0.61 (0.224)	0.13 (0.109)	8.57 (7.026)	–	-224.3	3.92
Gaussian	0.39 (0.071)	0.17 (0.044)	0.47 (0.175)	0.25 (0.098)	–	–	-225.7	3.92

10 Conclusions and extensions

This article has shown how DCS models with changing location and/or scale can be successfully extended to cover EGB2 conditional distributions. Most

CPI

inflation



$_1 0.pdf$

Figure 9: Filters for the level of UK inflation

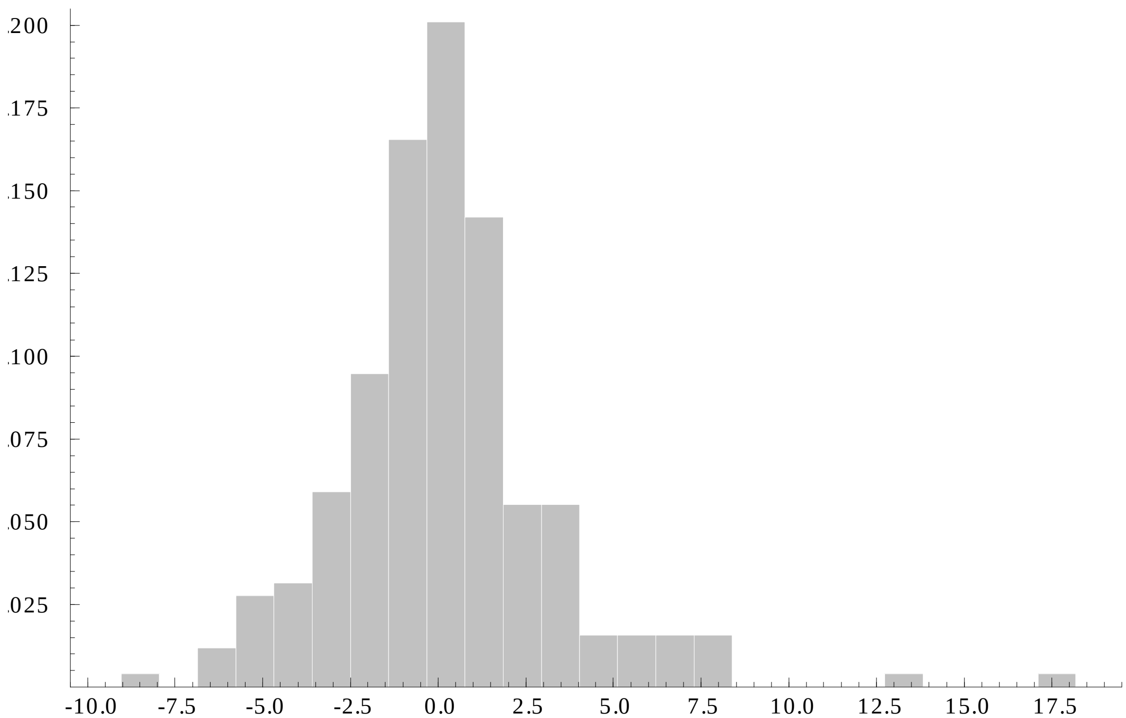


Figure 10: Histogram of residual from symmetric EGB2 model fitted to UK inflation.

of the theoretical results on the properties of DCS-t models, including the asymptotic distribution of ML estimators, carry over to EGB2 models. However, whereas the t-distribution has fat-tails, and hence subjects extreme observations to a form of soft trimming, the EGB2 distribution has light tails (but excess kurtosis) and hence gives a gentle form of Winsorizing. The examples show that the EGB2 distribution can give a better fit to some macroeconomic series and may even be appropriate for financial time series, such as exchange rates. For more complex modeling, using the DCS time series models to pre-adjust the data may be a more attractive alternative that arbitrarily trimming or Winsorizing the data.

The DCS models can be extended to include explanatory variables, thereby generalizing the nature and scope of dynamic regression. Extensions to handle multivariate series may be possible by following the approach in Yang et al (2011).

REFERENCES

Andres, P. and A. C. Harvey (2012). The dynamic location/scale model. Cambridge Working paper in Economics, CWPE 1240.

Asmussen, S. (2003). Applied Probability and Queues, Berlin, Springer Verlag.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *Annals of Mathematical Statistics* **25**, 573-578.

Creal, D., Koopman, S.J., and A. Lucas (2013). Generalized autoregressive score models with applications. J of Applied Econometrics (to appear).

Creal, D., Koopman, S.J. and A. Lucas (2011). A Dynamic Multivariate Heavy-Tailed Model for Time-Varying Volatilities and Correlations, Journal of Business and Economic Statistics, 29, 552-63.

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods.* 2nd edition. Oxford: Oxford Statistical Science Series.

Embrechts, P. , Kluppelberg, C. and T. Mikosch (1997). Modelling Extremal Events. Berlin: Springer Verlag.

Fernandez, C. and M. F. J. Steel (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association* **99**. 359-371.

Francq, C. and J-M Zakoïan (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications.* New York: John Wiley and Sons.

Giot, P. and S. Laurent (2003). Value at risk for long and short trading positions. *Journal of Applied Econometrics* **18**, 641-664.

Harvey A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails.* Econometric Society Monograph. Cambridge University Press.

Harvey, A. C. and A. Luati (2012). Filtering with Heavy Tails, CWPE 1255.

Harvey, A. C. and G. Sucarrat (2012). EGARCH Models with Fat Tails, Skewness and Leverage. Cambridge Working paper in Economics, CWPE 1236.

Jensen, S. T. and A. Rahbek (2004). Asymptotic inference for nonstationary GARCH. *Econometric Theory* **20**, 1203-26.

Kleiber, C. and S. Kotz (2003). *Statistical Size Distributions in Economics and Actuarial Sciences.* New York: Wiley

Koopman, S. J. Harvey, A. C., Doornik, J. A. and N. Shephard (2009). *STAMP 8.2 Structural Time Series Analysis Modeller and Predictor.* London: Timberlake Consultants Ltd.

Lange, K. L., Little, R. J. A. and M. G. Taylor (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association* **84**, 881-896

Lange, T., Rahbek, A. and S. T. Jensen (2010). Estimation and asymptotic inference in the AR-ARCH model. *Econometric Reviews* **30**, 129-53.

Lui, S., Mitchell, J. and M. Weale (2011) Qualitative business surveys: signal or noise? *J. R. Statist. Soc. A* (2011), 174, 327–348.

Maronna, R., Martin, D. and Yohai, V. (2006). *Robust Statistics: Theory and Methods*. John Wiley & Sons Ltd.

McDonald,. J.B. and S. B. White (1993): A comparison of some robust, adaptive, and partially adaptive estimators of regression models, *Econometric Reviews*, 12, 103-124

McDonald J. B. and Y. J. Xu (1995). A generalization of the beta distribution with applications. Journal of Econometrics 66: 133-152.

McNeil, A.J., Frey, R. and P.Embrechts (2005). Quantitative Risk Management. (Princeton University Press, Princeton)

Muler, N., Pena, D. and V. J. Yohai (2009). Robust estimation for ARMA models. *Annals of Statistics* **37**, 816-840.

Muler, N. and V. J. Yohai (2008). Robust estimates for GARCH models. *Journal of Statistical Planning and Inference* **138**, 2918-2940.

Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59, 347-370.

Prentice, R. L. (1975). Discrimination among some parametric models. *Biometrika*, 62, 607-614.

Wang, K-L, Fawson, C., Barrett, C.B. and J. B. McDonald.(2001). J*ournal of Applied Econometrics*, 16, 521-536.

Yang, X., Frees, E.W. and Z. Zhang (2011). A generalized beta copula with applications in modeling multivariate long-tailed data. *Insurance: Mathematics and Economics*, 49, 265-84.

Zhu, D. and J. W. Galbraith (2011). Modelling and forecasting expected shortfall with the generalised asymmetric student-t and asymmetric exponential power distributions. *Journal of Empirical Finance* **18**, 765-778.

Zhu, D. and V. Zinde-Walsh (2009). Properties and estimation of asymmetric exponential power distribution. *Journal of Econometrics*, 148, 86-99.

Appendix: Information matrix with the standard deviation parameterization

The information matrix for an EGB2 dynamic location model with the parameterization $\Psi = [\kappa, \phi, \omega, \lambda, \xi, \zeta]$, where $\lambda = \log(h/\nu)$ denotes the logarithm of the standard deviation, ν is the (inverse) scale parameter and $h \equiv \sqrt{\psi'(\xi) + \psi'(\zeta)}$ is as follows:

$$I \begin{pmatrix} \kappa \\ \phi \\ \omega \\ \sigma \\ \xi \\ \zeta \end{pmatrix} = \begin{bmatrix} \frac{h^2 e^{-2\lambda}\xi\zeta}{\xi+\zeta+1}\mathbf{D}(\boldsymbol{\kappa}, \boldsymbol{\phi}, \boldsymbol{\omega}) & \mathbf{d}I_{12} & \mathbf{d}I_{13} & \mathbf{d}I_{14} \\ \mathbf{d}'I_{21} & I_{22} & I_{23} & I_{24} \\ \mathbf{d}'I_{31} & I_{32} & I_{33} & I_{34} \\ \mathbf{d}'I_{41} & I_{42} & I_{43} & I_{44} \end{bmatrix}$$

where:

* $I_{12} = I_{21} = \frac{-he^{-\lambda}(\xi-\zeta-\xi\zeta(\psi(\xi)-\psi(\zeta)))}{\xi+\zeta+1}$.

$I_{13} = I_{31} = he^{-\lambda}\left[\frac{\varsigma}{\xi+\varsigma} + \frac{\eta_\xi}{\xi+\varsigma+1}\left(\xi - \varsigma - \xi\varsigma\left(\psi(\xi) - \psi(\varsigma)\right)\right)\right]$.

$I_{14} = I_{41} = he^{-\lambda}\left[-\frac{\xi}{\xi+\varsigma} + \frac{\eta_\varsigma}{\xi+\varsigma+1}\left(\xi - \varsigma - \xi\varsigma\left(\psi(\xi) - \psi(\varsigma)\right)\right)\right]$.

**$I_{22} = \frac{\xi\varsigma}{\xi+\varsigma+1}\left(\psi'(\xi) + \psi'(\varsigma) + \left(\psi(\xi) - \psi(\varsigma) - \frac{\xi-\varsigma}{\xi\varsigma}\right)^2 - \frac{\xi^2+\varsigma^2}{\xi^2\varsigma^2}\right) + 1$.

$I_{23} = I_{32} = -\left[\frac{1}{\xi+\varsigma} - \frac{\varsigma}{\xi+\varsigma}\left(\psi(\xi) - \psi(\varsigma)\right) + \eta_\xi I_{22}\right]$.

$I_{24} = I_{42} = -\left[\frac{1}{\xi+\varsigma} + \frac{\xi}{\xi+\varsigma}\left(\psi(\xi) - \psi(\varsigma)\right) + \eta_\varsigma I_{22}\right]$.

$I_{33} = \psi'(\xi) - \psi'\left(\xi+\varsigma\right) + \frac{2}{\xi+\varsigma}\eta_\xi - 2\eta_\xi\frac{\varsigma}{\xi+\varsigma}\left(\psi(\xi) - \psi(\varsigma)\right) + \eta_\xi^2 I_{22} = \psi'(\xi) - \psi'\left(\xi+\varsigma\right) + \frac{2\eta_\xi}{\xi+\varsigma}\left[1 - \varsigma\left(\psi(\xi) - \psi(\varsigma)\right)\right] + \eta_\xi^2 I_{22}$

$I_{34} = I_{43} = -\psi'\left(\xi+\varsigma\right) + \frac{\eta_\xi+\eta_\varsigma}{\xi+\varsigma} + \frac{1}{\xi+\varsigma}\left(\varsigma\eta_\varsigma - \xi\eta_\xi\right)\left(\psi(\xi) - \psi(\varsigma)\right) + \eta_\xi\eta_\varsigma I_{22}$.

$I_{44} = \psi'(\varsigma) - \psi'(\xi+\varsigma) + \frac{2}{\xi+\varsigma}\eta_\varsigma + 2\eta_\varsigma \frac{\xi}{\xi+\varsigma}(\psi(\xi) - \psi(\varsigma)) + \eta_\varsigma^2 I_{22}$.

$\eta_\xi \equiv \frac{\psi''(\xi)}{2h^2}$, $\eta_\varsigma \equiv \frac{\psi''(\varsigma)}{2h^2}$, and:

$$\mathbf{D}\begin{pmatrix}\kappa\\ \phi\\ \omega\end{pmatrix} = \begin{bmatrix} A & D & E\\ D & B & F\\ E & F & C\end{bmatrix}, \qquad \mathbf{d} = \begin{bmatrix} 0\\ 0\\ \frac{1-\phi}{1-a}\end{bmatrix}$$

with
$A = \sigma_u^2 = \sigma^2 h^2 \frac{\xi\varsigma}{\xi+\varsigma+1}$, $B = \frac{\kappa^2\sigma_u^2(1+a\phi)}{\left(1-\phi^2\right)(1-a\phi)}$, $C = \frac{(1-\phi)^2(1+a)}{1-a}$
$D = \frac{a\kappa\sigma_u^2}{1-a\phi}$, $E = \frac{c(1-\phi)}{1-a}$, $F = \frac{ac\kappa(1-\phi)}{(1-a)(1-a\phi)}$
$a = \phi - \kappa h^2 \frac{\xi\varsigma}{\xi+\varsigma+1}$
$b = \phi^2 - 2\phi\kappa h^2 \frac{\xi\varsigma}{\xi+\varsigma+1} + \kappa^2 \frac{h^4 \xi\varsigma(\xi+1)(\varsigma+1)(\xi+\varsigma)}{(\xi+\varsigma+3)(\xi+\varsigma+2)(\xi+\varsigma+1)}$
$c = \kappa\sigma h^3 \frac{\xi\zeta(\xi-\varsigma)}{(\xi+\varsigma+2)(\xi+\varsigma+1)}$.

In the symmetric case:

$$I\begin{pmatrix}\kappa\\ \phi\\ \omega\\ \lambda\\ \xi\end{pmatrix} = \begin{bmatrix} \frac{h^2 e^{-2\lambda}\xi^2}{2\xi+1}\mathbf{D}(\boldsymbol{\kappa},\boldsymbol{\phi},\boldsymbol{\omega}) & \mathbf{0} & \mathbf{0}\\ \mathbf{0}' & \frac{2\xi-2\xi^2\psi'(\xi)-1}{2\xi+1} & I_{23}\\ \mathbf{0}' & I_{32} & I_{33}\end{bmatrix}$$

$I_{23} = I_{32} = -\left(\frac{1}{\xi} + 2\eta_\xi \frac{2\xi-1}{2\xi+1} + 2\eta_\xi \frac{2\xi^2}{2\xi+1}\psi'(\xi)\right)$.
$I_{33} = \frac{4\eta_\xi}{\xi} + 4\eta_\xi^2 \frac{2\xi-1}{2\xi+1} + 4\eta_\xi^2 \frac{2\xi^2}{2\xi+1}\psi'(\xi) + 2\psi'(\xi) - 4\psi'(2\xi)$.

Note that, as when $\xi \to 0$, $\xi\eta_\xi \to -\frac{1}{2}$, both I_{23} and I_{33} go to ∞ as $\xi \to 0$.

With an exponential link function $\xi = \exp\left(\bar{\xi}\right)$, the information matrix becomes:

$$I\begin{pmatrix}\kappa\\ \phi\\ \omega\\ \lambda\\ \bar{\xi}\end{pmatrix} = \begin{bmatrix} \frac{h^2 e^{-2\lambda}\xi^2}{2\xi+1}\mathbf{D}(\boldsymbol{\kappa},\boldsymbol{\phi},\boldsymbol{\omega}) & \mathbf{0} & \mathbf{0}\\ \mathbf{0}' & \frac{2\xi-2\xi^2\psi'(\xi)-1}{2\xi+1} & I_{23}\\ \mathbf{0}' & I_{32} & I_{33}\end{bmatrix}$$

with

$$
\begin{aligned}
I_{23} &= I_{32} = -\left(1 + 2\xi\eta_\xi \tfrac{2\xi-1}{2\xi+1} + 2\xi\eta_\xi \tfrac{2\xi^2}{2\xi+1}\psi'\left(\xi\right)\right).\\
I_{33} &= 4\xi\eta_\xi + 4\left(\xi\eta_\xi\right)^2 \tfrac{2\xi-1}{2\xi+1} + 4\left(\xi\eta_\xi\right)^2 \tfrac{2\xi^2}{2\xi+1}\psi'\left(\xi\right) + \xi^2\left(2\psi'\left(\xi\right) - 4\psi'\left(2\xi\right)\right).
\end{aligned}
$$

When $\xi \to 0$ both I_{23} and I_{33} go to 0, so the matrix is singular, as in the scale parameterization.